DEFENCE R&D DÉFENSE

A Maximum Likelihood HF Direction Finding Estimator for High Latitude Distributed Signals

William Read
Defence Research Establishment Ottawa

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-085
April 1999

National Defence Défense nationale

Canada

DEFENCE R&D DÉFENSE

A Maximum Likelihood HF Direction Finding Estimator for High Latitude Distributed Signals

William Read
EW Sensors Group
Electronic Support Measures Section

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-085
April 1999

Project
5BD15

ABSTRACT

This report details the development and evaluation of a new direction finding estimator, the Spread Maximum Likelihood algorithm, targeted for high latitude HF propagation conditions. Of particular concern are patches of enhanced electron density and associated instabilities in the ionosphere which drift across the polar cap during darkness causing scattering of the propagated signal over a range of azimuth and elevation directions (the signals are spatially spread). The new estimator incorporates a spatial spreading model allowing it to simultaneously deal with both patch scattered signals and signals propagated by more normal propagation mechanisms, as well as to distinguish between the two types. Evaluation of the new estimator using both simulation and collected data show it to be considerably superior to both conventional and modern superresolution approaches for high latitude propagation conditions.

RÉSUMÉ

Ce rapport décrit une technique pour l'estimation de l'angle de gisement des ondes HF ainsi que son évaluation. L'algorithme intitulé "vraisemblence de l'étalement maximal" s'applique aux ondes HF transmisse dans les latitudes élevée. Il s'applique particulièrement aux zones de l'ionosphère où des perturbations associées à une densité d'électrons plus élevée se produisent. Ces zones dérivent au-dessus de la calotte polaire pendant la nuit, ce qui entraîne une diffusion du signal propagé sur une gamme d'azimuts et de directions d'élévation, ce qui veut dire que les signaux subissent un étalement spatial. La technique d'estimation comprend un modèle d'étalement spatial qui permet de traiter les signaux se propageant à traver les zones instables ainsi que les signaux propagés par des mécanismes plus conventionnels et de faire la distinction entre les deux. L'évaluation menée avec des données fictives et réelles a démontré que la méthode proposée est supérieur aux approches conventionnelles et modernes de superrésolution pour la propagation en haute latitude.

EXECUTIVE SUMMARY

The requirement exists to improve the accuracy of strategic high latitude HF direction finding (DF) systems. In the past, poor DF accuracy derived from Arctic measurements has led to low confidence in high latitude sites despite the strategic relevance of these sites for transmitter geolocation.

The biggest problem appears to be patches of enhanced electron density and associated instabilities in the F layer of the ionosphere which drift across the polar cap during darkness. These patches scatter the signal over a range of azimuth and elevation directions (the signals are spatially spread) which are very different from the expected bearing given the transmitter's location.

Although the bearings of these spatially spread signals provide no indication of the true bearing of the transmitter, there is evidence that sporadic-E propagation (i.e. reflections from the E layer of the ionosphere) may sometimes be supported during these conditions and that this propagation mode yields the desired signal bearing. Unfortunately, conventional DF approaches estimate the bearing of the most dominant propagation mode which will generally be the patch scattered signals. Newer superresolution DF algorithms, which can simultaneously DF signals from several directions, tend to become overloaded by the range of directions occupied by the scattered signals. Hence there is a need for a new approach which is capable of dealing with this unique high-latitude phenomena.

In this report, a new high latitude HF DF algorithm is introduced. The theoretical development of this new algorithm is based on maximum likelihood principles and the key feature is that a model of the spatial spreading is incorporated. The performance of this new algorithm – the spread maximum likelihood (SML) method – is evaluated using both simulated data and off-air data collected at the Arctic site CFS Alert. This evaluation included the effects of noise, the amount of spatial spreading, and the shape of the spread region. It also included an evaluation of the ability to detect a weaker point-source (no spatial spreading) signal in the presence of stronger spatially spread signals. The results of this evaluation show that the new algorithm is not only capable of dealing with spatially spread signals but also provides information on the amount of spreading. This extra information is invaluable since it provides a useful means of qualifying signal accuracy. For example, a large amount of spatial spreading indicates a patch scattered signal which can be ignored, while little spreading indicates the signal was propagated

by a more conventional propagation mode and more likely yields the correct transmitter bearing.

As is already evident, the main advantages of the SML algorithm are that, for high latitude HF conditions, it provides more useful information about the received signals, and it is considerably more accurate than both conventional and modern superresolution DF approaches when propagation conditions are poor (i.e., when patch scattering occurs). For more benign conditions, the performance of the SML algorithm is at least as good as the other approaches. For these reasons, serious consideration should be given to its implementation at high latitude DF sites.

The main disadvantage of the SML algorithm is that it is computationally very intensive which makes it an order of magnitude or more slower than other approaches. For operational purposes, this problem could be overcome by using the SML algorithm only when patch scattering conditions are suspected, and using faster DF methods otherwise.

Future research should focus on developing realtime implementations of the SML algorithm. This is a possibility which could be realizable within the next few years given the ever increasing speed of modern computing power.

TABLE OF CONTENTS

	Page
ABSTRACT/RÉSUMÉ	iii
EXECUTIVE SUMMARY	v
TABLE OF CONTENTS	vii
LIST OF FIGURES	ix
LIST OF TABLES	xi
1.0 INTRODUCTION	1
2.0 SPREAD MAXIMUM LIKELIHOOD METHOD OVERVIEW	3
2.1 Maximum Likelihood Estimation	3
2.2 The Cost Function	5
2.3 Modeling the Signal Environment	6
3.0 MODEL PARAMETER ESTIMATION	13
3.1 Initial Noise Estimate	15
3.2 Azimuth-Elevation Search	16
4.0 PARAMETER REFINEMENT	19
4.1 Noise Power Gradient	22
4.2 Signal Power Gradient	23
4.3 Bearing Gradient	23
4.4 Spread Gradient	24
4.4.1 The Uniform Rectangular Grid	25
4.4.2 The Raised-Cosine Elliptical Grid	31
4.5 Out-of-Bounds Signals	36
4.5.1 The Uniform Rectangular Grid	36
4.5.2 The Raised-Cosine Elliptical Grid	37
5.0 SML ALGORITHM SUMMARY	42
5.1 Main SML Routine	42
5.2 Gradient Ascent Routine	45
6.0 SML CONTROL PARAMETER SETTINGS	49
6.1 Signal Model Grid Spacing	52
6.2 Data Block Size	54
6.3 Taper Function	57
6.4 Assumed Number of Signals	60
7.0 SML PERFORMANCE	63
7.1 Effect of Noise	63
7.2 Effect of Signal Spreading	65

7.3 Detection of a Weaker Signal 67
7.4 Performance using High Latitude Off-Air Data 70

8.0 CONCLUSIONS AND RECOMMENDATIONS 80

REFERENCES 82

LIST OF FIGURES

		Page
Figure 1:	Grid of point-source signals used to represent a spread signal	11
Figure 2:	Geometry of the uniform rectangular grid	27
Figure 3:	Geometry of the raised-cosine elliptical grid	32
Figure 4:	Position of point-source signals in grid for narrow spread signals . .	34
Figure 5:	Position of point-source signals in grid involving out-of-bounds signals	37
Figure 6:	Antenna array geometry	50
Figure 7:	Antenna array beamwidth as a function elevation	50
Figure 8:	Effect of grid spacing on SML estimation performance	53
Figure 9:	Effect of number of data samples on SML estimation performance .	55
Figure 10:	Effect of number of data samples on estimation of spread signals . .	56
Figure 11:	Asymmetric power distribution used for spread signal simulations .	59
Figure 12:	Effect of assumed number of signals on SML estimation performance	61
Figure 13:	Effect of noise on SML estimation performance	64
Figure 14:	Effect of signal spread on SML accuracy	65
Figure 15:	Effect of angular spacing on the ability to detect a weaker point-source signal in the presence of a stronger spread-source signal	69
Figure 16:	MUSIC azimuth bearing estimates for 100 trials	70
Figure 17:	Detection of a weaker point-source signal in the presence of a stronger spread-source signal using MUSIC	71
Figure 18:	Antenna geometry of the Vortex array	72
Figure 19:	MUSIC results for high latitude off-air data	73
Figure 20:	MUSIC results for high latitude off-air data showing dominant direction	74

Figure 21: SML bearing results for high latitude off-air data 75

Figure 22: SML results for high latitude off-air data showing dominant direction 77

Figure 23: SML spreading results for the dominant signal 78

Figure 24: SML results for noise and dominant signal power 79

LIST OF TABLES

		Page
Table 1:	Simulation Signal Parameters	49
Table 2:	Simulation Signal Parameters for Uniform Rectangular and Asymmetric Spread Distributions	58
Table 3:	Effect of Taper Function on SML Performance	59
Table 4:	Signal Parameters for Signal Detectability Simulation	67

1.0 INTRODUCTION

The requirement exists to improve the accuracy of strategic HF direction finding systems, particularly in the Arctic. In the past, poor DF accuracy derived from Arctic measurements has led to low confidence in high latitude sites despite the strategic relevance of these sites for transmitter geolocation.

The biggest problem appears to be patches of enhanced electron density and associated instabilities in the F layer of the ionosphere which drift across the polar cap during darkness in a roughly antisunward direction at speeds ranging from a few hundreds to over one thousand meters per second [1]. These patches can cause scattering of a signal from azimuth directions which are very different from the true bearing of the transmitter.

One avenue of investigation being pursued is the development of new DF algorithms which are better matched to the high latitude HF signal environment. Currently popular DF algorithms assume that the incoming signal can be modeled as a point source, or equivalently, the incoming signal has a planar wavefront. This is reasonable if the size of the transmission source is extremely small relative to its range; the size of the DF array is also small relative to the transmitter range; the ionosphere acts as a perfect or near-perfect reflector; and local site multipath can be ignored. Unfortunately at high latitudes, during periods where scattering from large moving patches is occurring, the received signals are distributed over a range of bearings in both azimuth and elevation. Current DF algorithms handle this phenomena by splitting it up into several signals coming from the same area of the ionosphere.

Although representing a signal scattered from a region of the ionosphere (called a spread-source signal in this report) by several point-source signals is still informative, it degrades the ability of the algorithm to detect weaker signals that may also be present. For example, previous studies have indicated that the E layer of the ionosphere acts as a good reflector when it supports signal propagation (i.e. sporadic-E) between the transmitter and receiver. Hence, it is important that a DF algorithm be able to detect and determine the bearing of a weaker sporadic-E reflected signal in the presence of one or more stronger spread signals.

In attempts to meet these requirements, a number of DF algorithms were developed at McMaster University [2]. Although these algorithms showed promise, in their development a strong effort was made to keep the processing considerations reasonable and this,

unfortunately, resulted in algorithms which were only able to deal with a single signal.

To overcome the single signal limitation, processing considerations were abandoned in favour of developing the most accurate approach possible capable of dealing with multiple spread signals. To this end, the maximum likelihood approach was chosen, since this approach leads to optimal estimators when all known information about the signal environment is taken into account. The result is a new estimator called the *Spread Maximum Likelihood Estimator* whose development and testing is the subject of the rest of this report.

The layout of the rest of the report is as follows. In Section 2, the maximum likelihood method is introduced followed by the development of the specific signal model used by the spread maximum likelihood (SML) variant. In Sections 3-5, a detailed description of the SML algorithm is provided, including the initial value assessment in Section 3, the refinement procedure in Section 4, and the step-by-step algorithm summary in Section 5. In Section 6, the optimum settings for various algorithm parameters is investigated through simulation. Comparisons are also made with the MUSIC DF algorithm (an algorithm that is commonly used in advanced DF processing) using simulated data and off-air data collected with a 12-channel DF system in the Arctic. Finally, in Section 7 the conclusions and recommendations are presented.

2.0 SPREAD MAXIMUM LIKELIHOOD METHOD OVERVIEW

2.1 Maximum Likelihood Estimation

A successful, albeit often computationally intensive, approach to estimation is based on the maximum likelihood method. Essentially the idea is to find the most likely state of a signal process given a set of measurement observations made of the process. Since this is a statistical approach, the method applies to cases where either the signal generation is a random process, or the measurements have been corrupted by additive noise (which is a random process), or both.

Assuming the random processes are all normally Gaussian distributed, and measurements are made using N sensors, then the associated probability density function is given by [3]

$$f(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{K-1}) = \frac{1}{[\pi^N \det \mathbf{C}]^K} e^{-\mathrm{trace}((\mathbf{X}-\mathbf{M})^H \mathbf{C}^{-1}(\mathbf{X}-\mathbf{M}))} \tag{1}$$

where the superscript H denotes the conjugate-transpose operation, the vectors $\mathbf{x}_0, ..., \mathbf{x}_{K-1}$ represents the random complex measurement data associated with all N sensors for time instances $t = t_0, t_1, ..., t_{K-1}$ as defined by

$$\mathbf{x}_k = \begin{bmatrix} x_0(k) \\ x_1(k) \\ x_2(k) \\ \vdots \\ x_{N-1}(k) \end{bmatrix} \quad \text{for } 0 \leq k < K. \tag{2}$$

Additionally, the matrix $\mathbf{X}$ is merely a compact way of representing all K measurement vectors and is given by

$$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{K-1}], \tag{3}$$

the matrix $\mathbf{M}$ represents the corresponding mean values of the measurements, and $\mathbf{C}$ is the $N \times N$ covariance matrix describing the correlations among sensors. The exact definitions of the matrices $\mathbf{M}$ and $\mathbf{C}$ depend on how the above density function is applied to the particular estimation problem to be solved, as will be seen. Once this equation has been properly set up, the maximum likelihood solution is then found by maximizing the probability density function by making the appropriate choice of $\mathbf{M}$ and/or $\mathbf{C}$ from among the allowable choices. The optimum choices of $\mathbf{M}$ and/or $\mathbf{C}$ can then be related

back to the signal parameters of interest.

For direction finding purposes $\mathbf{x}_1, ..., \mathbf{x}_{K-1}$ are taken to represent the complex baseband outputs from an array of N antennas. The definitions for $\mathbf{M}$ and $\mathbf{C}$ depends on the assumptions made about the signals. The two most commonly used assumptions are either

1. the signal is deterministic but unknown, or,
2. the signal is stochastic with a zero-mean Gaussian distribution.

Choosing the first assumption, the data can be decomposed as

$$x_n(k) = m_n(k) + \eta_n(k) \tag{4}$$

where $m_n(k)$ is the exact (but unknown) signal value and $\eta_n(k)$ is the value of the noise corrupting the measurement. From this expression it follows that the elements of $\mathbf{M}$ are given by $m_n(k)$, and the elements of $\mathbf{C}$ then represent the covariance of the noise processes only, as defined by

$$c_{mn} = E[\eta_m(k)\eta_n(k)^*] \quad \text{for} \quad 0 \leq k < K \text{ and } 0 \leq m, n < N. \tag{5}$$

The approach based on the first assumption is called the *deterministic maximum likelihood method.*

Choosing the second signal assumption leads to

$$m_n(k) = 0 \tag{6}$$

and

$$c_{mn} = E[x_m(k)x_n(k)^*] \quad \text{for} \quad 0 \leq k < K \text{ and } 0 \leq m, n < N \tag{7}$$

since the sum of two or more zero-mean Gaussian processes is still a zero-mean Gaussian process. The approach based on the stochastic assumption is called the *stochastic maximum likelihood method.*

Comparing these two approaches reveals that the main differences are the constraints placed on the estimated complex signal amplitude(s). The deterministic signal assumption (usually) leads to no constraint while the stochastic signal assumption leads to a Gaussian

distribution. Given that man-made signals are bounded (finite power) and often have a Gaussian-like distribution, the second assumption generally results in better estimator performance [4] (see also [5] for a comparison of the deterministic approach with better modeled methods). An exception to this might be the case of constant modulus signals (e.g., fm, psk, fsk, etc.) since a constant amplitude constraint could more easily be added to the deterministic approach than the stochastic approach. However, for the HF signals considered in this report, the dynamic nature of the ionosphere tends to destroy the constant modulus properties of the received signal, so that the stochastic maximum likelihood approach was followed (in fact the deterministic approach was tried but was not found to be nearly as successful).

2.2 The Cost Function

In this section, the stochastic maximum likelihood method is further refined. The central key to this refinement is the development of a signal plus noise model which accurately reflects the high latitude HF environment. This involves accounting for signals which are diffusely reflected or scattered off a highly disturbed region of the ionosphere with an angular extent of several degrees or more in both azimuth and elevation, hence the name *Spread Maximum Likelihood Estimator*. By comparison, the standard stochastic maximum likelihood approach considers only point-sources signals.

Before beginning the model development, the probability density function (1) can be simplified using (6) to give

$$f(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{K-1}) = \frac{1}{[\pi^N \det \mathbf{C}]^K} e^{-\text{trace}(\mathbf{X}^H \mathbf{C}^{-1} \mathbf{X})} \tag{8}$$

by using the stochastic signal model. The objective remains the same as before, that is, find the unknown covariance matrix $\mathbf{C}$ which maximizes this expression. This is equivalent to maximizing the cost function $L_c = \ln f(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{K-1})$ or,

$$L_c = -NK \ln(\pi) - K \ln(\det \mathbf{C}) - \text{trace}(\mathbf{X}^H \mathbf{C}^{-1} \mathbf{X}). \tag{9}$$

Since the addition or multiplication by a constant value has no effect on the maximization, the cost function can be simplified to

$$L = -\ln(\det \mathbf{C}) - \text{trace}(\mathbf{R}\mathbf{C}^{-1}) \tag{10}$$

where $\mathbf{R}$ is the *data covariance matrix* constructed from $\mathbf{X}$ using

$$\mathbf{R} = \frac{1}{K}\mathbf{X}\mathbf{X}^H. \tag{11}$$

2.3 Modeling the Signal Environment

The modeling aspect comes into play when $\mathbf{C}$ is being selected. A model is used to generate $\mathbf{C}$ based on input modeling parameters such as, for example, the number of signals, signal bearings, signal amplitudes/powers, and noise powers. For this reason, $\mathbf{C}$ is referred to as the *model covariance matrix* in this report. Once the model has been setup, a common approach, and the approach employed here, is to choose some initial model parameters, generate $\mathbf{C}$, and then determine the cost function value L. The model parameters are then successively refined and $\mathbf{C}$ recomputed until the maximum cost function value has been achieved. The model parameter values corresponding to this maximum value are then taken to be the optimum or maximum likelihood estimates.

The particular model used to generate the model covariance matrix depends on many factors including the transmitter(s) and receiver characteristics, the signal propagation environment, the noise sources, and so on. One way to set this model up, is to consider the generation of synthetic data which imitates the collected data, and then use this synthetic data to determine $\mathbf{C}$ according to

$$\mathbf{C} = \frac{1}{K}\mathbf{Y}\mathbf{Y}^H \tag{12}$$

where $\mathbf{Y}$ is the matrix of model data and has the same form and dimensions as $\mathbf{X}$.

Proceeding in this manner, it becomes necessary to make certain assumptions about the environment in order to simplify the derivations. Initially, the following ideal assumptions can be made:

1. M incoming point-source signals arrive from M different directions.
2. The noise is additive and Gaussian.
3. The receiver channels are linear and perfectly matched.
4. The receive antennas are isotropic and located at the same height.
5. There is no coupling among the receive antennas or with the local environment.

Modifications to these assumptions will then be introduced as appropriate.

Based on these simplifying assumptions, the received signal can be decomposed as

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + ... + \mathbf{Y}_M + \mathbf{N} \tag{13}$$

where the matrices $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_M$ represent the model data for the M individual signals, and $\mathbf{N}$ is the modeled noise. The matrix $\mathbf{Y}_m$, for $0 < m \leq M$, can be defined in vector form as

$$\mathbf{Y}_m = \mathbf{e}_m \mathbf{a}_m^H \tag{14}$$

where $\mathbf{e}_m$ is the array response (or steering) vector for the m^{th} signal, and $\mathbf{a}_m$ is the corresponding signal amplitude vector. The definitions for the elements of the steering vector are given by

$$\mathbf{e}_m = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j\frac{2\pi}{\lambda}(x_0 \sin\phi_m \cos\psi_m + y_0 \cos\phi_m \cos\psi_m)} \\ e^{j\frac{2\pi}{\lambda}(x_1 \sin\phi_m \cos\psi_m + y_1 \cos\phi_m \cos\psi_m)} \\ \vdots \\ e^{j\frac{2\pi}{\lambda}(x_{N-1} \sin\phi_m \cos\psi_m + y_{N-1} \cos\phi_m \cos\psi_m)} \end{bmatrix} \tag{15}$$

where λ is the signal wavelength, x_n and y_n are the Cartesian coordinates for antenna n (with the phase center of the array located at the origin), ϕ_m is the azimuth angle of the m^{th} signal measured clockwise with respect to the Y-axis of the coordinate system, and ψ_m is the elevation angle measured with respect to the X-Y plane (the ground). The definition for the elements of the signal amplitude vector is given by

$$\mathbf{a}_m = \begin{bmatrix} a_m(0)^* \\ a_m(1)^* \\ \vdots \\ a_m(K-1)^* \end{bmatrix} \tag{16}$$

where $a_m(0), a_m(1), ..., a_m(K-1)$ are the received complex amplitudes of the m^{th} signal for time instances $t_0, t_1, ..., t_{K-1}$.

Analytically, the preceding approach to generating the model covariance matrix is useful, however, it requires too many model parameters. For example, the generation of $\mathbf{Y}$ requires selecting the azimuth angle, elevation angle, and complex amplitudes for each signal plus the corresponding complex noise values for each sensor. With the exception

of the angles, this must be repeated for each time instance resulting in a total number of $2M + 2(M + N)K$ real model parameters (remembering that one complex parameter is equivalent to two real parameters). Additionally the noise and signal amplitudes must also be chosen so that statistically they have complex Gaussian distributions.

A more practical approach is to rewrite the model covariance as a sum of the noise covariances, the signal covariances, and the signal cross-covariances, or

$$\mathbf{C} = \sigma^2 \mathbf{C}_\eta + \sum_{m=1}^{M} \mathbf{C}_{mm} + \sum_{m=1}^{M-1} \sum_{n=m+1}^{M} (\mathbf{Q}_{mn} + \mathbf{Q}_{mn}^H) \tag{17}$$

where σ^2 is the noise power, $\mathbf{C}_\eta$ is the normalized noise covariance matrix (trace $\mathbf{C}_\eta = 1$), $\mathbf{C}_{mm}$ is the signal covariance matrix for the m^{th} signal, and $\mathbf{Q}_{mn}$ is the signal cross-covariance matrix. The generation of these matrices is discussed in the following paragraphs.

The noise covariance matrix $\mathbf{C}_\eta$ is assumed to be known a priori and will not be considered as part of the estimation process. The determination of $\mathbf{C}_\eta$ can be done either through theoretical statistical considerations, or through measurements. For example, if the noise is known to be white Gaussian in nature with equal but uncorrelated amplitudes in each channel, then $E[\eta_m(k)\eta_m(k)^*] = \sigma^2$ and $E[\eta_m(k)\eta_n(k)^*] = 0$ for $0 \leq m, n < N$ and $m \neq n$, hence

$$\mathbf{C}_\eta = \frac{1}{N}\mathbf{I}_N \tag{18}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Alternatively, if data measurements can be taken when no signals are present, then

$$\mathbf{C}_\eta = \frac{\mathbf{X}\mathbf{X}^H}{\text{trace}(\mathbf{X}\mathbf{X}^H)}. \tag{19}$$

More elaborate procedures could be developed involving a number of measurement sets with the same noise environment but different signal directions, however the development of this kind of approach is beyond the scope of this report. It suffices to say that joint estimation of both the noise and signal characteristics should be avoided if possible since it leads to poorer accuracy.

The signal covariance matrix $\mathbf{C}_{mm}$ can be defined in terms of the model data as

$$\mathbf{C}_{mm} = \frac{1}{K}\mathbf{Y}_m\mathbf{Y}_m^H \tag{20}$$

which can be further expanded in terms of the component amplitude and steering vectors as

$$\mathbf{C}_{mm} = \frac{\mathbf{a}_m^H \mathbf{a}_m}{K} \mathbf{e}_m \mathbf{e}_m^H. \tag{21}$$

In a similar fashion, the cross-covariance matrix $\mathbf{Q}_{mn}$ can also be defined and expanded to get

$$\begin{aligned} \mathbf{Q}_{mn} &= \frac{1}{K} \mathbf{Y}_m \mathbf{Y}_n^H \\ &= \frac{\mathbf{a}_m^H \mathbf{a}_n}{K} \mathbf{e}_m \mathbf{e}_n^H \qquad \text{for } n > m. \end{aligned} \tag{22}$$

At this point, the number of model parameters has not been changed (except for the noise covariances which are now assumed to be known). Simplifications can be introduced to the expressions for $\mathbf{C}_{mm}$ and $\mathbf{Q}_{mn}$ by recognizing that the term $\mathbf{a}_m^H \mathbf{a}_m / K$ is the average signal power over the given time interval and can be replaced by the single parameter s_m. Equation (21) can then be rewritten as

$$\mathbf{C}_{mm} = s_m \mathbf{e}_m \mathbf{e}_m^H. \tag{23}$$

Additionally, defining the complex correlation coefficient

$$\rho_{mn} = \frac{\mathbf{a}_m^H \mathbf{a}_n}{|\mathbf{a}_m||\mathbf{a}_n|} \tag{24}$$

where $|\mathbf{a}| = (\mathbf{a}^H \mathbf{a})^{\frac{1}{2}}$, then $\mathbf{Q}_{mn}$ can be rewritten as

$$\mathbf{Q}_{mn} = \rho_{mn} (s_m s_n)^{\frac{1}{2}} \mathbf{e}_m \mathbf{e}_n^H. \tag{25}$$

Hence the K complex amplitude model parameters for each signal have been replaced by a single power parameter plus up to $M-1$ complex correlation coefficients. Also counting the bearing angles required for each signal (ϕ_m and ψ_m), and the noise power, then the number of real model parameters is $3M + M(M-1) + 1$ which is a considerable reduction in parameters when $K \gg M$ (which will normally be true in practice).

A further major reduction in the number of parameters is possible if the correlations ρ_{mn} are known. For the HF skywave environment, signals from different transmitters will be uncorrelated. Signals originating from the same transmitter but reflected off different layers of the ionosphere will also be uncorrelated due to the large path length differences

usually encountered (i.e., the path delay time differences will be greater than the inverse bandwidth of the signal). Hence one can assume that

$$\rho_{mn} = 0 \qquad \text{for } 0 < m < n \leq M \tag{26}$$

which eliminates the third term in (17) and reduces the number of real model parameters to $3M + 1$.

A violation of the assumption of uncorrelated signals is the case of local site multipath. However, like the problem of determining the noise correlations, including the signal correlations in the estimation process is highly undesirable (although in the standard stochastic ML approach this is done [6]). Hence it is assumed that either the receiver site is well chosen, or signal correlations can be determined independently and corrected as done, for example, in [7].

Up until this point, only incoming point-source signals have been considered. For HF, this assumes perfect reflections off the ionosphere, which is a reasonable assumption for benign atmospheric conditions. At high latitudes, however, the ionosphere can sometimes become highly disturbed [8] resulting in behaviour which is not consistent with an ideal reflector. For example, measurements have shown that during these disturbed conditions Doppler shifts of up to 45 Hz or more can occur in conjunction with large eigenvalue spreading in the measured data covariance matrix [9]. These measurements suggest that the transmitter signal is scattered from highly dynamic regions of the ionosphere yielding the large Doppler spreads. These measurements also suggest that the regions are sufficiently wide (angular spreads of several degrees or more) that signal decorrelation occurs across the region causing the large eigenvalue spreading.

This interpretation is also supported by HF propagation research work carried out at CRC [10]. This work suggests that the most likely mechanism for the ionospheric disturbances is moving patches of enhanced electron density with sizes greater than several kilometers which drift across the polar cap during darkness. As these patches progress, they give rise to short-lived small-scale electron irregularities on the scale of less than one kilometer. These irregularities become very elongated along the earth's magnetic field lines and tend to scatter incident radio waves through 360° in azimuth with little or no scattering in elevation. Given the random and short-lived nature of these irregularities, the patch acts as a temporal and spatial incoherent radio scattering region.

The interpretation of conditions encountered at high latitudes leads to a modified

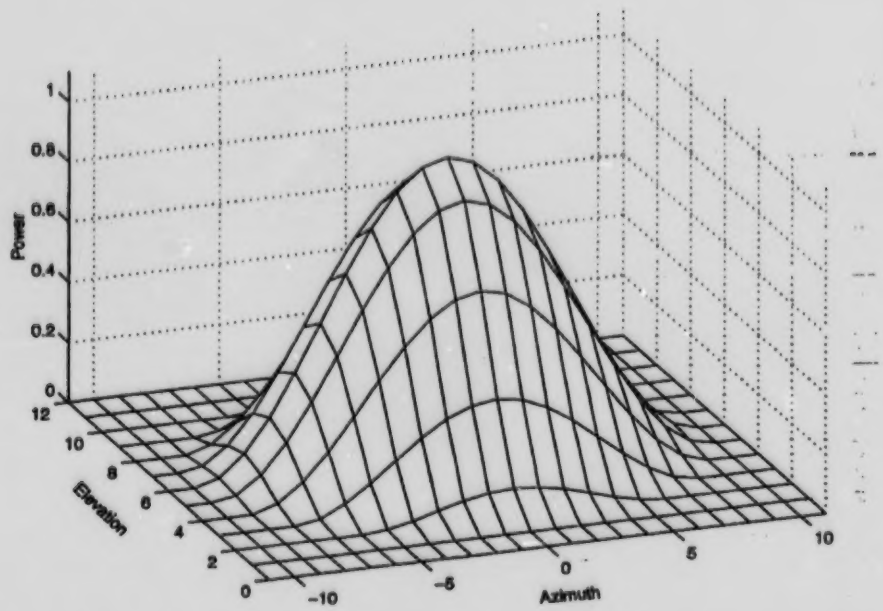


Figure 1: Grid of point-source signals used to represent a spread signal. Each grid node represents location and power of each corresponding source signal.

description for the signal covariance matrix $\mathbf{C}_{mm}$. To begin with, the scattering region is divided into a grid where each section in the grid is made small enough that it can be accurately represented by a single perfectly reflected point-source signal. Figure 1 shows an example plotting the power of the point-source signals versus azimuth and elevation angle for a single region. The power towards the edges of the modeled region is tapered according to some predetermined taper function which is chosen to best represent the actual environment. Given that signals scattered from different parts of the same region are uncorrelated, the model covariance matrix can be generated according to the sum of the individual covariances of the component point-source signals, or

$$\mathbf{C}_{mm} = s_m \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) \mathbf{e}_{mi} \mathbf{e}_{mi}^H \tag{27}$$

where M_g is the number of component point-source signals in the grid, $\mathbf{e}_{mi}$ represents the steering vectors of the grid signals, $\beta_m(\Phi, \Psi)$ is the taper function which is positive

real-valued and normalized so that

$$\sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) \;=\; 1 \tag{28}$$

and Φ_{mi} and Ψ_{mi} are the normalized grid angles. The normalized grid angles are defined by

$$\Phi_{mi} = \frac{\phi_m - \phi_{mi}}{\Delta_{\phi_m}} \tag{29}$$

$$\Psi_{mi} = \frac{\psi_m - \psi_{mi}}{\Delta_{\psi_m}} \tag{30}$$

where ϕ_m and ψ_m are the respective azimuth and elevation bearing angles of the center of the grid, ϕ_{mi} and ψ_{mi} are the respective azimuth and elevation angles of the component point-source signals, and Δ_{ϕ_m} and Δ_{ψ_m} are the signal spread parameters defining the angular extent of the spread signal (width of the grid) in azimuth and elevation, respectively.

In order to be mathematically usable, the taper function must vary smoothly as a function of the azimuth and elevation spread parameters, and the differentiated taper function must be defined for all valid values of the spread parameters. For example, in a rectangular grid with uniform fixed spacings and equal powers for all individual point-source signals, the taper function changes in discrete steps as each new column/row of point-source signals is introduced or removed when the azimuth/elevation spread changes sufficiently. The choice of taper function and grid layout (since this affects the taper function) is discussed in Section 4.4.

Using (17), (27) and incorporating the simplifications to the model covariance generation already described, then

$$\mathbf{C} = \sigma^2 \mathbf{C}_\eta + \sum_{m=1}^{M} s_m \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) \mathbf{e}_{mi} \mathbf{e}_{mi}^H \tag{31}$$

where the noise power σ^2, signal power s_m, azimuth bearing ϕ_m, elevation bearing ψ_m, azimuth spread Δ_{ϕ_m}, and elevation spread Δ_{ψ_m} are the only input values which are explicitly adjusted in order to maximize the cost function L.

3.0 MODEL PARAMETER ESTIMATION

The previous section examined the generation of the model covariance matrix and its evaluation via the cost function. In this section, an overview of the model parameter estimation procedure is provided along with details on determining the initial model parameter estimates. Details on the parameter refinement procedure are provided later in Section 4.0.

A major problem with determining the optimum model parameters in maximum likelihood problems is that it is usually difficult or impossible to develop a closed-form solution. The standard approach is to employ some form of search procedure to find the solution. An exhaustive search of all possible model parameter values could be performed to do this, however, with more than a very small number of model parameters, this is a prohibitively time-consuming process. Faster approaches such as gradient ascent/descent techniques are useful, but require the initial model parameter to be reasonably close to the optimum solution, otherwise the search may converge on a false solution. An approach which has been used successfully in the past, and works reasonably well with uncorrelated signals, is the Alternating Projection Maximization (APM) method [11]. Although heavily modified, the approach employed here is conceptually similar.

The basic procedure is to build up the model covariance matrix one signal at a time. In the very first stage an estimate of the noise power is made assuming the data consists of noise only (Section 3.1). Using this as the starting model covariance matrix, a single point-source signal is added. This is done by sweeping the signal bearing parameters through 360° in azimuth and 90° in elevation, while keeping the noise power fixed and optimizing the signal power (Section 3.2). Several potential signal parameter solutions corresponding to higher values of the cost function L are selected. For each of these candidate solutions, a modified gradient ascent technique is employed to jointly refine both the noise power and all the signal parameters including spread (see Section 4.0). The refined parameter set yielding the highest value of the cost function is then retained and all other refined candidate sets are thrown out. This completes stage one.

In the next stage, using the new model covariance matrix corresponding to the refined model parameter set from the previous iteration, a new set of point-source signal candidates are identified using the azimuth and elevation search method in the same manner as in the first stage. This is followed by another joint refinement of all signal and noise parameters, and again, only the refined parameter set yielding the highest value of L is

retained. This completes stage two.

The process described for stage two is repeated again and again until the last (M^{th}) signal has been introduced and the model parameters have been refined. The resultant model parameters are then taken to be the optimum maximum likelihood estimates, the determination of which will be presented later.

An extra step is added to the joint refinement procedure when the bearing of a candidate point-source signal falls within the angular spread region of one or more of the refined signals selected in the previous stage. In this case, the refinement procedure involving the candidate point-source signal is repeated with the following special modifications to the parameters of the affected signals (including the candidate signal and the refined signals with which it "collides", but no others): the azimuth and elevation spread values are set to zero, and the signal powers set equal to the average power of the colliding signals. If this modified refinement yields the highest value of L, then the new refined parameter set is retained, otherwise it is rejected.

Performing the azimuth-elevation search at each step using a point-source signal tends to favour the detection of point-source signals over spread-source signals. However, since point-source signals are of the most interest (i.e., for high latitude HF DF applications, a point-source signal is more likely to yield the true bearing of the transmitter than a real spread-source signal), this enhancement is not a bad thing.

The identification of several candidate model parameter sets during the search and the ensuing refinement to identify the best solution is a modification to the original APM method in order to improve the probability of finding the correct solution.

The modified gradient ascent technique used for parameter refinement is also a deviation from the original APM method. It was used because it was found to converge faster than both the technique proposed for the APM method and the standard gradient ascent/descent method.

The next few sections provide a detailed discussion of the initial noise estimate, the azimuth-elevation search procedure for the introduction of new signals, and the modified gradient ascent technique used to refine the signal and noise parameters.

3.1 Initial Noise Estimate

The initial noise estimate is made by assuming that there are no signals present and then finding the value of σ^2 which maximizes the cost function L. The no signal assumption is equivalent to assuming

$$\mathbf{C} = \sigma^2 \mathbf{C}_\eta. \tag{32}$$

Substituting this into the expression for cost function (10) gives

$$\begin{aligned} L_0 &= -\ln(\det(\sigma^2 \mathbf{C}_\eta)) - \text{trace}(\mathbf{R}(\sigma^2 \mathbf{C}_\eta)^{-1}) \\ &= -N\ln(\sigma^2) - \ln(\det \mathbf{C}_\eta) - \frac{1}{\sigma^2}\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1}) \end{aligned} \tag{33}$$

where L_0 denotes the cost function value for zero assumed signals and N is the number of sensors.

The cost function can be maximized (or minimized) by differentiating L_0 with respect to σ^2 and then finding the value of σ^2 which sets the differentiated expression to zero. Hence differentiating yields

$$\frac{\partial L_0}{\partial \nu} = \frac{-N}{\sigma^2} + \frac{1}{\sigma^4}\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1}) \tag{34}$$

where $\nu = \sigma^2$, and then setting the result to zero and solving gives

$$\sigma^2 = \frac{1}{N}\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1}). \tag{35}$$

For this value of σ^2 the value of the second derivative $\partial^2 L_0/\partial \nu^2$ is

$$\frac{-N^3}{\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1})^2} < 0 \tag{36}$$

indicating that value of σ^2 given by (35) produces a maximum (not minimum) in the cost function L_0.

The expression for the optimum value of σ^2 (in the no signal case) can be inserted back into the expression for the cost function L_0 (10) to give a new expression independent of σ^2 (but where the optimum value of σ^2 is implicitly assumed). The result is

$$L_0 = N\ln(N) - N - \ln(\det \mathbf{C}_\eta) - N\ln(\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1})). \tag{37}$$

3.2 Azimuth-Elevation Search

In the azimuth-elevation search, a new point-source signal is added to the model covariance matrix. In this section, the search procedure is developed so that the azimuth and elevation angles of the new signal can be adjusted without having to explicitly adjust the signal power as well. Additionally, although the model noise covariance $\mathbf{C}_\eta$ is required in this search, an initial estimate of the noise power is not.

Beginning with the known model covariance represented by $\mathbf{C}_\alpha$, the addition of a new point-source signal yields

$$\mathbf{C} = \mathbf{C}_\alpha + s_m \mathbf{e}_m \mathbf{e}_m^H \tag{38}$$

where the indexing begins with $m = 1$ and is incremented after each stage until $m = M$. The inverse of the matrix $\mathbf{C}$ can be found using the augmented matrix inversion lemma to give

$$\mathbf{C}^{-1} = \mathbf{C}_\alpha^{-1} - \frac{s_m \mathbf{C}_\alpha^{-1} \mathbf{e}_m \mathbf{e}_m^H \mathbf{C}_\alpha^{-1}}{1 + s_m \mathbf{e}_m^H \mathbf{C}_\alpha^{-1} \mathbf{e}_m} \tag{39}$$

where it is also assumed here, and throughout the rest of this report, that $\mathbf{C}$ and $\mathbf{C}_\alpha$ are positive definite so that $\mathbf{C}^{-1}$ and $\mathbf{C}_\alpha^{-1}$ exist and $\mathbf{e}_m^H \mathbf{C}_\alpha^{-1} \mathbf{e}_m > 0$. For simplicity a new vector

$$\mathbf{q}_m = \mathbf{C}_\alpha^{-1} \mathbf{e}_m \tag{40}$$

is introduced so that the expression for $\mathbf{C}^{-1}$ becomes

$$\mathbf{C}^{-1} = \mathbf{C}_\alpha^{-1} - \frac{s_m \mathbf{q}_m \mathbf{q}_m^H}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m} \tag{41}$$

Plugging the two expressions for $\mathbf{C}$ and $\mathbf{C}^{-1}$ into (10) and simplifying

$$\begin{aligned} L_m &= -\ln(\det(\mathbf{C}_\alpha + s_m \mathbf{e}_m \mathbf{e}_m^H)) - \operatorname{trace}(\mathbf{R}(\mathbf{C}_\alpha^{-1} - \frac{s_m \mathbf{q}_m \mathbf{q}_m^H}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m})) \\ &= -\ln(\det(\mathbf{I} + s_m \mathbf{q}_m \mathbf{e}_m^H) \det(\mathbf{C}_\alpha)) - \operatorname{trace}(\mathbf{R}\mathbf{C}_\alpha^{-1}) + \operatorname{trace}\left(\frac{s_m \mathbf{R} \mathbf{q}_m \mathbf{q}_m^H}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m}\right) \\ &= -\ln((1 + s_m \mathbf{e}_m^H \mathbf{q}_m) \det(\mathbf{C}_\alpha)) - \operatorname{trace}(\mathbf{R}\mathbf{C}_\alpha^{-1}) + \left(\frac{s_m \mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m}\right) \end{aligned} \tag{42}$$

where L_m is the cost function after the m^{th} signal is introduced into the model.

To reduce the search to only the azimuth and elevation angles, the cost function can

be differentiated with respect to the signal power and then solved to find the optimum value of the signal power. Differentiation with respect to s_m produces

$$\frac{\partial L_m}{\partial s_m} = \frac{-\mathbf{e}_m^H \mathbf{q}_m}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m} + \frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{1 + s_m \mathbf{e}_m^H \mathbf{q}_m} - \frac{s_m \mathbf{e}_m^H \mathbf{q}_m \mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{(1 + s_m \mathbf{e}_m^H \mathbf{q}_m)^2} \tag{43}$$

and setting this result to zero then solving for s_m gives

$$s_m = \frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m - \mathbf{e}_m^H \mathbf{q}_m}{(\mathbf{e}_m^H \mathbf{q}_m)^2}. \tag{44}$$

The second derivative $\partial^2 L_m / \partial s_m^2$ for this value of s_m is

$$\frac{-(\mathbf{e}_m^H \mathbf{q}_m)^4}{(\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m)^2} < 0 \tag{45}$$

indicating that this value of s_m is the desired maximum value, and not a minimum.

Finally, the expression for s_m can be reinserted into (42), and after some manipulation the result is given by

$$\begin{aligned} L_m &= -\ln(\det \mathbf{C}_\alpha) - \mathrm{trace}(\mathbf{R}\mathbf{C}_\alpha^{-1}) - \ln\left(\frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{\mathbf{e}_m^H \mathbf{q}_m}\right) + \frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{\mathbf{e}_m^H \mathbf{q}_m} - 1 \\ &= L_{m-1} - \ln\left(\frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{\mathbf{e}_m^H \mathbf{q}_m}\right) + \frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{\mathbf{e}_m^H \mathbf{q}_m} - 1 \end{aligned} \tag{46}$$

which is the cost function evaluated for the optimum signal power. Note that in the above expression, L_{m-1} is the maximum cost function value from the previous stage.

Using (46), the search for the optimum signal parameters can be carried out by varying the azimuth and elevation angles only. The maximization process can be simplified even more by noticing that the constant terms in (46) have no effect on the maximization, and that maximizing $g(x) - \ln g(x)$ with respect to x is the same as maximizing $g(x)$, where $g(x)$ is a positive real-valued function. Hence maximizing L_m is the same as maximizing

$$S_m(\phi_m, \psi_m) = \frac{\mathbf{q}_m^H \mathbf{R} \mathbf{q}_m}{\mathbf{e}_m^H \mathbf{q}_m} \tag{47}$$

or

$$S_m(\phi_m, \psi_m) = \frac{\mathbf{e}_m^H \mathbf{C}_\alpha^{-1} \mathbf{R} \mathbf{C}_\alpha^{-1} \mathbf{e}_m}{\mathbf{e}_m^H \mathbf{C}_\alpha^{-1} \mathbf{e}_m}. \tag{48}$$

The function $S_m(\phi_m, \psi_m)$ is not only the specialized cost function for the single signal case (or additive signal case), but it can also be interpreted as the whitened beamformer spectrum. To see this, consider that in estimation problems a common practice is to filter the data in such a way as to whiten the noise spectrum (this assumes that the noise spectrum is known a priori). Once this has been done, optimal estimators designed for white noise can be used. This simplifies the problem of having to design different estimators for white noise and coloured noise.

For the estimator problem at hand, $\mathbf{C}_a$ can be treated as the known "noise" covariance, and assuming it to be full rank, then there will exist an appropriate $N \times N$ spatial whitening filter $\mathbf{W}$ such that

$$\mathbf{W}\mathbf{W}^H = \mathbf{C}_a^{-1} \tag{49}$$

(how $\mathbf{W}$ is actually determined is unimportant here). Using this filter, the whitened data is formed according to

$$\mathbf{X}_w = \mathbf{W}^H\mathbf{X} \tag{50}$$

which leads to the whitened covariance matrix

$$\mathbf{R}_w = \mathbf{W}^H\mathbf{R}\mathbf{W}. \tag{51}$$

To account for the affect of the whitening filter on the response of the sensor array, the steering vector $\mathbf{e}$ is modified in the same way as the data and then normalized to give

$$\mathbf{e}_w = \frac{\mathbf{W}^H\mathbf{e}}{\sqrt{\mathbf{e}^H\mathbf{W}\mathbf{W}^H\mathbf{e}}} = \frac{\mathbf{W}^H\mathbf{e}}{\sqrt{\mathbf{e}^H\mathbf{C}_a^{-1}\mathbf{e}}}. \tag{52}$$

The whitened beamformer spectrum is then given by

$$S_m(\phi_m, \psi_m) = \mathbf{e}_w^H\mathbf{R}_w\mathbf{e}_w \tag{53}$$

which becomes (48) if $\mathbf{R}_w$, and $\mathbf{e}_w$ are expanded in terms of (50)-(52).

From this discussion, it can be seen that the introduction of a new point-source signal into the covariance model can be accomplished using a simple beamformer to search in azimuth and elevation. Signal power does not have to be explicitly tested. Additionally, the first signal can also be determined without knowledge of the noise power since the maximum of $S_m(\phi_m, \psi_m)$ remains unchanged whether $\mathbf{C}_a = \sigma^2\mathbf{C}_\eta$ is used or $\mathbf{C}_a = \mathbf{C}_\eta$ is used.

4.0 PARAMETER REFINEMENT

In the previous section, initialization of the model parameters was discussed. In this section, the gradient ascent method is introduced and developed to further refine and optimize the model parameter estimates. The principle strategy employed is to incrementally adjust the values of the model parameters according to the local gradient of the cost function until a maximum in the cost function is found (i.e., climbs the "hill" until the peak is reached). Since this approach finds the local maximum, care must be taken in the initial model parameter estimates to ensure that the local maximum corresponds to the global maximum.

Using α to represent any of the real-valued model parameters (e.g., σ^2, or s_m, or ϕ_m, etc.), then the gradient ascent parameter adjustment at each step i is performed according to

$$\alpha_i = \alpha_{i-1} + \mu_\alpha G(\alpha_{i-1}) \tag{54}$$

where $G(\alpha_i)$ is the slope or gradient with respect to the model parameter, and μ_α is a positive real value used to control the step size. The above expression is performed simultaneously for all the adjustable model parameters before proceeding to the next step, that is, model parameters for step i are computed using only the model parameters from step $i-1$.

The parameter adjustment step size μ is also modified at each step (in this discussion the subscript α is used only when discussing the step size for a single model parameter, and dropped when discussing the step sizes for all the model parameters). In each step, after all the model parameters have been adjusted, the new parameter set is tested to see if the cost function is improved. If $L_i > L_{i+1}$ the parameter adjustment is considered successful and the step size for all parameters is slightly increased ($\mu \rightarrow 1.2\mu$). If $L_i \leq L_{i+1}$, the adjustment is considered unsuccessful, the new parameter set is discarded, and the step size for all parameters is decreased ($\mu \rightarrow \mu/3$).

In addition to these modifications to the step size, a special modification is also used to improve the behaviour of the gradient ascent technique when various kinds of structures in the cost function are encountered. For these structures, a few parameters will be close to their optimum values but their corresponding slopes on both sides of these optimum values will be large, while for other parameters their optimum values are further away and the slopes are smaller (e.g., consider a long ridge which slowly rises along the length of the ridge but has very steep sides). The result is that the closer parameter values will

dominate the cost function (the larger slope means a greater change in the cost function for a given increment), and the values of μ_α will need to be reduced accordingly to keep these parameter values in the vicinity of their optimum values. With a small value of μ_α, the parameters that are further away take a large number of steps before they come close to their optimum values.

One way to detect this slow convergence condition is to examine the behaviour of the gradient function. For a parameter close to its optimum value, the gradient will oscillate between being positive and negative as the adjustments overshoot back and forth across the optimum value ($G(\alpha_{i-1})G(\alpha_{i-2}) < 0$). If this condition is detected after the new parameter set has been deemed successful, then the step size μ_α is replaced by $\mu_\alpha/3$.

The definition of the gradient function depends on the model parameter being adjusted, however before getting into the finer details, some generalizations can be made. The gradient of the cost function (10) is given by

$$\begin{aligned} G(\alpha) &= \frac{\partial L}{\partial \alpha} \\ &= -\frac{\partial \ln(\det \mathbf{C})}{\partial \alpha} - \frac{\partial \mathrm{trace}(\mathbf{R}\mathbf{C}^{-1})}{\partial \alpha}. \end{aligned} \tag{55}$$

This can be simplified somewhat by considering each of the two terms in the above expression individually, as will be seen.

Beginning with the first term, consider the following eigen-decomposition of the model covariance matrix

$$\mathbf{C} = \sum_{i=1}^{N} \lambda_i \mathbf{v}_i \mathbf{v}_i^H \tag{56}$$

where λ_i represents the eigenvalues and $\mathbf{v}_i$ represents the eigenvectors of $\mathbf{C}$ with the properties $\mathbf{v}_i^H \mathbf{v}_i = 1$ and $\mathbf{v}_i^H \mathbf{v}_k = 0$ for $i \neq k$. Since the determinant of a matrix is equal to the product of it's eigenvalues, the first term in (55) becomes

$$\frac{\partial \ln(\det \mathbf{C})}{\partial \alpha} = \frac{\partial \sum_{i=1}^{N} \ln(\lambda_i)}{\partial \alpha}. \tag{57}$$

Performing the indicated differentiation and utilizing the eigenvectors and their properties

to expand and rearrange the expression, then

$$\begin{aligned}
\frac{\partial \ln(\det \mathbf{C})}{\partial \alpha} &= \sum_{i=1}^{N} \lambda_i^{-1} \frac{\partial \lambda_i}{\partial \alpha} \\
&= \sum_{i=1}^{N} \lambda_i^{-1} \frac{\partial \lambda_i}{\partial \alpha} + \frac{\partial \mathbf{v}_i^H \mathbf{v}_i}{\partial \alpha} \qquad \left(\text{where} \quad \frac{\partial \mathbf{v}_i^H \mathbf{v}_i}{\partial \alpha} = 0 \quad \text{since} \quad \mathbf{v}_i^H \mathbf{v}_i = 1\right) \\
&= \sum_{i=1}^{N} \lambda_i^{-1} \frac{\partial \lambda_i}{\partial \alpha} + \frac{\partial \mathbf{v}_i^H}{\partial \alpha} \mathbf{v}_i + \mathbf{v}_i^H \frac{\partial \mathbf{v}_i}{\partial \alpha} \\
&= \sum_{i=1}^{N} \lambda_i^{-1} (\lambda_i \mathbf{v}_i^H \frac{\partial \mathbf{v}_i}{\partial \alpha} + \frac{\partial \lambda_i}{\partial \alpha} + \lambda_i \frac{\partial \mathbf{v}_i^H}{\partial \alpha} \mathbf{v}_i) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i^{-1} \mathbf{v}_i^H (\frac{\partial \mathbf{v}_k}{\partial \alpha} \lambda_k \mathbf{v}_k^H + \mathbf{v}_k \frac{\partial \lambda_k}{\partial \alpha} \mathbf{v}_k^H + \mathbf{v}_k \lambda_k \frac{\partial \mathbf{v}_k^H}{\partial \alpha}) \mathbf{v}_i \\
&= \text{trace}(\sum_{i=1}^{N} \sum_{k=1}^{N} \mathbf{v}_i \lambda_i^{-1} \mathbf{v}_i^H (\frac{\partial \mathbf{v}_k}{\partial \alpha} \lambda_k \mathbf{v}_k^H + \mathbf{v}_k \frac{\partial \lambda_k}{\partial \alpha} \mathbf{v}_k^H + \mathbf{v}_k \lambda_k \frac{\partial \mathbf{v}_k^H}{\partial \alpha})) \\
&= \text{trace}(\sum_{i=1}^{N} \mathbf{v}_i \lambda_i^{-1} \mathbf{v}_i^H \sum_{k=1}^{N} \frac{\partial \mathbf{v}_k}{\partial \alpha} \lambda_k \mathbf{v}_k^H + \mathbf{v}_k \frac{\partial \lambda_k}{\partial \alpha} \mathbf{v}_k^H + \mathbf{v}_k \lambda_k \frac{\partial \mathbf{v}_k^H}{\partial \alpha}) \\
&= \text{trace}(\sum_{i=1}^{N} \mathbf{v}_i \lambda_i^{-1} \mathbf{v}_i^H \sum_{k=1}^{N} \frac{\partial \mathbf{v}_k \lambda_k \mathbf{v}_k^H}{\partial \alpha}) \\
&= \text{trace}(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \alpha}). \qquad (58)
\end{aligned}$$

The second term can be simplified by first differentiating the equality $\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$, which gives

$$\frac{\partial \mathbf{C}\mathbf{C}^{-1}}{\partial \alpha} = \mathbf{0} \qquad (59)$$

where $\mathbf{0}$ is a matrix of zeros having the same dimensions as $\mathbf{C}$. Expanding the lefthand side

$$\frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{C}^{-1} + \mathbf{C} \frac{\partial \mathbf{C}^{-1}}{\partial \alpha} = \mathbf{0} \qquad (60)$$

and rearranging

$$\frac{\partial \mathbf{C}^{-1}}{\partial \alpha} = -\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \alpha} \mathbf{C}^{-1}. \qquad (61)$$

Using this relationship, and carrying out the indicated differentiation of the second term

in (55), then

$$\begin{aligned}\frac{\partial \text{trace}(\mathbf{R}\mathbf{C}^{-1})}{\partial \alpha} &= \text{trace}(\mathbf{R}\frac{\partial \mathbf{C}^{-1}}{\partial \alpha}) \\ &= -\text{trace}(\mathbf{R}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha}\mathbf{C}^{-1}) \\ &= -\text{trace}(\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha}). \end{aligned} \tag{62}$$

Finally, a general expression for the gradient can be found by substituting (58) and (62) back into (57), which gives

$$\begin{aligned} G(\alpha) &= -\text{trace}(\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha}) + \text{trace}(\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha}) \\ &= \text{trace}\left((\mathbf{C}^{-1}\mathbf{R} - \mathbf{I})\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha}\right). \end{aligned} \tag{63}$$

Detailed derivations of the gradient as a function of noise power, signal power, signal bearing, and bearing spread are given in the following sections. To simplify the expressions that result, the definition

$$\mathbf{P} = (\mathbf{C}^{-1}\mathbf{R} - \mathbf{I})\mathbf{C}^{-1} \tag{64}$$

is used. The gradient is then given by

$$G(\alpha) = \text{trace}(\mathbf{P}\frac{\partial \mathbf{C}}{\partial \alpha}). \tag{65}$$

Additionally, the definition for the model covariance $\mathbf{C}$ used in the following derivations is given by (31) and repeated here as

$$\mathbf{C} = \sigma^2\mathbf{C}_\eta + \sum_{m=1}^{M} s_m \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi})\mathbf{e}_{mi}\mathbf{e}_{mi}^H. \tag{66}$$

4.1 Noise Power Gradient

Letting $\alpha = \sigma^2$ and differentiating the model covariance matrix expression with respect to the noise power gives

$$\frac{\partial \mathbf{C}}{\partial \alpha} = \mathbf{C}_\eta, \tag{67}$$

and incorporating this in (65), the noise power gradient becomes

$$G(\sigma^2) = \text{trace}(\mathbf{P}\mathbf{C}_\eta). \tag{68}$$

4.2 Signal Power Gradient

Differentiating the model covariance matrix expression with respect to the power of the m^{th} signal gives,

$$\frac{\partial \mathbf{C}}{\partial s_m} = \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) \mathbf{e}_{mi} \mathbf{e}_{mi}^H, \tag{69}$$

and using $\alpha = s_m$ in (65), then the signal power gradient becomes

$$G(s_m) = \text{trace}(\mathbf{P} \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) \mathbf{e}_{mi} \mathbf{e}_{mi}^H). \tag{70}$$

4.3 Bearing Gradient

Differentiating the model covariance matrix expression with respect to the azimuth bearing angle of the m^{th} signal gives

$$\frac{\partial \mathbf{C}}{\partial \phi_m} = s_m \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi}) (\frac{\partial \mathbf{e}_{mi}}{\partial \phi_m} \mathbf{e}_{mi}^H + \mathbf{e}_{mi} \frac{\partial \mathbf{e}_{mi}^H}{\partial \phi_m}). \tag{71}$$

Using the definition for the steering vector (15) and the fact that ϕ_{mi} differs from ϕ_m by a predetermined constant offset, then

$$\frac{\partial \mathbf{e}_{mi}}{\partial \phi_m} = \mathbf{A}_{mi} \mathbf{e}_{mi} \tag{72}$$

where $\mathbf{A}_{mi}$ is a diagonal matrix defined by

$$\text{diag}(\mathbf{A}_{mi}) = j \frac{2\pi}{\lambda} \begin{bmatrix} x_0 \cos\phi_{mi} \cos\psi_{mi} - y_0 \sin\phi_{mi} \cos\psi_{mi} \\ x_1 \cos\phi_{mi} \cos\psi_{mi} - y_1 \sin\phi_{mi} \cos\psi_{mi} \\ \vdots \\ x_{N-1} \cos\phi_{mi} \cos\psi_{mi} - y_{N-1} \sin\phi_{mi} \cos\psi_{mi} \end{bmatrix}. \tag{73}$$

Using these definitions and $\alpha = \phi_m$ in (65), the azimuth bearing gradient becomes

$$G(\phi_m) = s_m \text{trace}(\mathbf{P} \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi})(\mathbf{A}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{A}_{mi})). \tag{74}$$

Similarly, the elevation bearing gradient can be derived to get

$$G(\psi_m) = s_m \text{trace}(\mathbf{P} \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi})(\mathbf{B}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{B}_{mi})) \tag{75}$$

where $\mathbf{B}_{mi}$ is a diagonal matrix defined by

$$\text{diag}(\mathbf{B}_{mi}) = -j\frac{2\pi}{\lambda}\begin{bmatrix} x_0 \sin\phi_{mi}\sin\psi_{mi} + y_0\cos\phi_{mi}\sin\psi_{mi} \\ x_1 \sin\phi_{mi}\sin\psi_{mi} + y_1\cos\phi_{mi}\sin\psi_{mi} \\ \vdots \\ x_{N-1}\sin\phi_{mi}\sin\psi_{mi} + y_{N-1}\cos\phi_{mi}\sin\psi_{mi} \end{bmatrix}. \tag{76}$$

A special modification to the elevation bearing gradient occurs if part of the spread region is cut-off by the horizon or 90° vertical. The expression is then represented by

$$\begin{aligned} G(\psi_m) \;=\; s_m \text{trace} \Bigg(& \mathbf{P} \sum_{i=1}^{M_g} \frac{\partial \beta_m(\Phi_{mi}, \Psi_{mi})}{\partial \psi_m} \mathbf{e}_{mi}\mathbf{e}_{mi}^H \\ & + \mathbf{P} \sum_{i=1}^{M_g} \beta_m(\Phi_{mi}, \Psi_{mi})(\mathbf{B}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{B}_{mi}) \Bigg). \end{aligned} \tag{77}$$

Since the complete derivation requires the exact definition of the taper function, this derivation is discussed separately in Section 4.5.

4.4 Spread Gradient

Differentiating the model covariance with respect to the azimuth or elevation spread of the m^{th} signal yields

$$\begin{aligned} \frac{\partial \mathbf{C}}{\partial \Delta_m} \;=\; s_m \sum_{i=1}^{M_g} \Bigg(& \frac{\partial \beta_m(\Phi_{mi}, \Psi_{mi})}{\partial \Delta_m} \mathbf{e}_{mi}\mathbf{e}_{mi}^H \\ & + \beta_m(\Phi_{mi}, \Psi_{mi}) \frac{\partial \mathbf{e}_{mi}}{\partial \Delta_m} \mathbf{e}_{mi}^H \;+\; \beta_m(\Phi_{mi}, \Psi_{mi}) \mathbf{e}_{mi} \frac{\partial \mathbf{e}_{mi}^H}{\partial \Delta_m} \Bigg) \end{aligned} \tag{78}$$

where Δ_m represents either Δ_{ϕ_m} or Δ_{ψ_m}. Defining $\mathbf{C}_{mi}$ as the model covariance matrix for each grid signal, or

$$\mathbf{C}_{mi} \doteq \beta_m(\Phi_{mi}, \Psi_{mi})\mathbf{e}_{mi}\mathbf{e}_{mi}^H, \tag{79}$$

then the derivative of the model covariance matrix can also be represented in a more compact form as

$$\frac{\partial \mathbf{C}}{\partial \Delta_m} = s_m \sum_{i=1}^{M_g} \frac{\partial \mathbf{C}_{mi}}{\partial \Delta_m}. \tag{80}$$

The corresponding form for the gradient is given by

$$G(\Delta_m) = s_m \text{trace}(\mathbf{P} \sum_{i=1}^{M_g} \frac{\partial \mathbf{C}_{mi}}{\partial \Delta_m}). \tag{81}$$

Without further defining the grid layout and taper functions it is not possible to go any farther, so for this report, two different choices of grid shape/taper functions were investigated. They were as follows:

1. A rectangular shaped grid using uniform grid spacing except on the outside edge, and a uniform power distribution for the individual point-source signals.
2. An elliptical shaped grid with uniform grid spacing and a raised-cosine power distribution for the individual point-source signals.

Although the second choice was expected to more closely resemble the real world, assessing both choices provided a means of comparing performance when very different patch shapes were assumed for the same data.

In the following, the spread gradient is defined for both taper functions under the assumption that neither the estimated signal elevation bearing nor spread angles result in grid signals below the horizon or above vertical. The special case when this assumption is not true is considered in the Section 4.5.

4.4.1 The Uniform Rectangular Grid

Figure 2 shows both the general shape of a rectangular grid and the arrangement of point-source signals in one row of the grid (the arrangement in a column is the same). With the exception of the two outside signals, the grid signal locations are uniformly spaced

with uniform power and the central signal located at (ϕ_m, ψ_m). The outside signals have their power levels scaled according to the angular distance to the neighbouring signal in order to provide a smooth transition when the spread is varied to the point that new rows/columns of point-source signals must be added or removed from the grid.

The determination of the values of the taper function for the various grid signals begins with the calculation of the numbers of these signals as a function of the spread parameters. Along each row (defined by constant elevation with increasing azimuth angles measured left to right), but excluding the outermost or nonuniformly spaced grid signals, there will be

$$M_{\phi_m} = 2\,\mathrm{int}(\frac{\Delta_{\phi_m}}{2d_{\phi_m}}) + 1 \tag{82}$$

grid positions where d_{ϕ_m} is the angular grid spacing within the row (it is also the column spacing). Similarly, along each column (defined by constant azimuth with increasing elevation measured bottom to top), there will be

$$M_{\psi_m} = 2\,\mathrm{int}(\frac{\Delta_{\psi_m}}{2d_{\psi_m}}) + 1 \tag{83}$$

grid positions, where d_{ψ_m} is the spacing within the column (it is also the row spacing). Ideally d_{ϕ_m} and d_{ψ_m} are chosen to be as small as possible, but since this increases the number of computations required, reasonable choices are one fifth the array 3 dB beamwidth (which is usually different in the azimuth and the elevation directions). The spacing requirement is discussed and evaluated in more detail in Section 6.1.

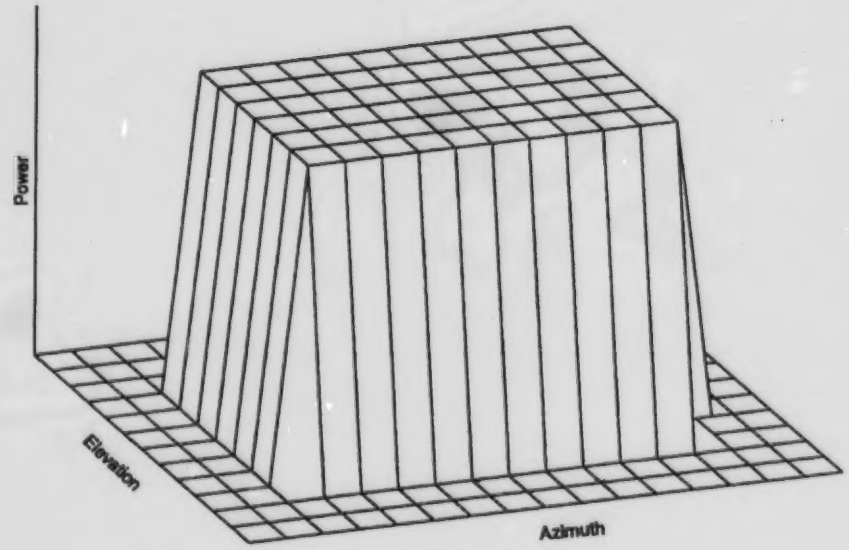
The $M_{\phi_m} \times M_{\psi_m}$ grid signals in the interior of the grid will have constant power, and the corresponding value of the taper function can be defined as β_{0m}. The value of the taper function for the grid signals at the edges is a function of the spacing at the edge relative to the spacing of the interior signals. Along the rows, the edge spacing is given by

$$d_a = \frac{1}{2}(\Delta_{\phi_m} - (M_{\phi_m} - 1)d_{\phi_m}). \tag{84}$$

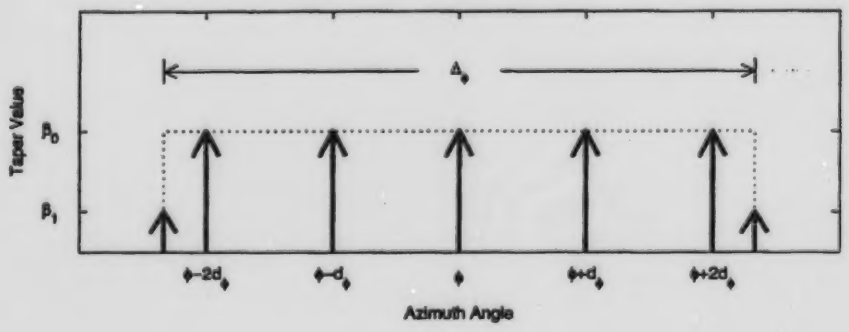
Similarly, along the columns the edge spacing is given by

$$d_e = \frac{1}{2}(\Delta_{\psi_m} - (M_{\psi_m} - 1)d_{\psi_m}). \tag{85}$$

Excluding the corner signals of the grid, the corresponding value of the taper function for



(a)



(b)

Figure 2: Geometry of the uniform rectangular grid showing (a) the general shape, and (b) an example of the position and power of the point-source signals making up one row of a grid.

the edges of each row is given by

$$\beta_{1m} = \beta_{0m}\frac{d_a}{d_{\phi_m}} = \frac{\beta_{0m}}{2}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} - M_{\phi_m} + 1\right) \tag{86}$$

and of each column is given by

$$\beta_{2m} = \beta_{0m}\frac{d_e}{d_{\psi_m}} = \frac{\beta_{0m}}{2}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} - M_{\psi_m} + 1\right). \tag{87}$$

For the four corners of the grid, the values of the taper function will be a function of both the row and column edge spacing values leading to

$$\beta_{3m} = \beta_{0m}\frac{d_a d_e}{d_{\phi_m} d_{\psi_m}} = \frac{\beta_{0m}}{4}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} - M_{\psi_m} + 1\right)\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} - M_{\phi_m} + 1\right). \tag{88}$$

Since the sum of the values of the taper function for all the grid positions is equal to one, then

$$M_{\phi_m} M_{\psi_m} \beta_{0m} + 2M_{\psi_m}\beta_{1m} + 2M_{\phi_m}\beta_{2m} + 4\beta_{3m} = 1. \tag{89}$$

Substituting for β_{1m}, β_{2m}, and β_{3m} using (86)-(88), and then rearranging in terms of β_{0m} the expression becomes

$$\begin{aligned}\beta_{0m} &= \Bigg(M_{\phi_m} M_{\psi_m} + M_{\psi_m}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} - M_{\phi_m} + 1\right) \\ &\quad + M_{\phi_m}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} - M_{\psi_m} + 1\right) + \left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} - M_{\phi_m} + 1\right)\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} - M_{\psi_m} + 1\right)\Bigg)^{-1} \\ &= \left(\frac{\Delta_{\phi_m}\Delta_{\psi_m}}{d_{\phi_m} d_{\psi_m}} + \frac{\Delta_{\phi_m}}{d_{\phi_m}} + \frac{\Delta_{\psi_m}}{d_{\psi_m}} + 1\right)^{-1}.\end{aligned} \tag{90}$$

Having now defined the shape of the grid and the taper function, it is now possible to continue the derivation of the gradient from (78). For signals in the interior of the grid, only the power varies as a function of the azimuth or elevation spread, so that $\partial \mathbf{C}_{mi}/\partial\Delta_m$ becomes

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_m} = \frac{\partial \beta_{0_m}}{\partial \Delta_m}\mathbf{e}_{mi}\mathbf{e}_{mi}^H \tag{91}$$

where β_{0_m} is the value of β_{0m} calculated for the m^{th} grid signal. Using the result from

(90) and carrying out the derivative operation then

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \frac{-\beta_{0_m}^2}{d_{\phi_m}}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} + 1\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H. \tag{92}$$

Excluding the corner signals, the signals on the top and bottom edges of the grid also vary only in power as the azimuth spread varies. Noting that the power is given by β_{2_m}, which can be expanded in terms of β_{0_m} using (87), then for the top and bottom edge signals the appropriate expression is

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \frac{-\beta_{0_m}\beta_{2_m}}{d_{\phi_m}}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} + 1\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H \tag{93}$$

which is similar to (92).

For signals on the left and right edges of the grid (and again excluding the corner signals), both the power and signal direction vary according to the azimuth spread. The relationship between azimuth spread and signal direction for the right and left edge signals is given by

$$\phi_{mi} = \phi_m \pm \frac{\Delta_{\phi_m}}{2} \tag{94}$$

where "+" is used for the right edge and "–" is used for the left edge. Using this relationship plus the fact that the power level is given by β_{1_m}, and partially differentiating $\mathbf{C}_{mi}$ with respect to Δ_{ϕ_m}, then

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \frac{\partial \beta_{1_m}}{\partial \Delta_{\phi_m}}\mathbf{e}_{mi}\mathbf{e}_{mi}^H \pm \frac{\beta_{1_m}}{2}\left(\mathbf{A}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{A}_{mi}\right) \tag{95}$$

where $\mathbf{A}$ was previously defined in Section 4.3. Using the definition for β_{1m} given in (86) and the definition for β_{0m} given in (90) in order to carry out the rest of the differentiation, the above result becomes

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \frac{\beta_{0m}}{2d_{\phi_m}}\left(1 - 2\beta_{1m}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} + 1\right)\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H \pm \frac{\beta_{1_m}}{2}\left(\mathbf{A}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{A}_{mi}\right). \tag{96}$$

Finally, for the signals at the corners of the grid, the differentiation proceeds in a similar manner except that the power level is given by β_{3m} which is defined in terms of

β_{0m} by (88). The resultant equation is given by

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \left(\frac{\beta_{2m}}{2d_{\phi_m}} - \frac{\beta_{0m}\beta_{3m}}{d_{\phi_m}}\left(\frac{\Delta_{\psi_m}}{d_{\psi_m}} + 1\right)\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H \pm \frac{\beta_{3m}}{2}(\mathbf{A}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{A}_{mi}) \quad (97)$$

where, as before, "+" is used for the corners on the right edge of the grid and "–" is used for corners on the left edge of the grid.

Using the same procedure to differentiate the covariance matrix with respect to the elevation spread, then for signals located in the interior of the grid

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\psi_m}} = \frac{-\beta_{0m}^2}{d_{\psi_m}}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} + 1\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H. \quad (98)$$

Excluding the grid signals at the corners, then for signals on the right and left edges

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\psi_m}} = \frac{-\beta_{0m}\beta_{1m}}{d_{\psi_m}}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} + 1\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H, \quad (99)$$

and for signals on the top and bottom edges

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\psi_m}} = \frac{\beta_{0m}}{2d_{\psi_m}}\left(1 - 2\beta_{2m}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} + 1\right)\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H \pm \frac{\beta_{2m}}{2}(\mathbf{B}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{B}_{mi}) \quad (100)$$

where

$$\psi_{mi} = \psi_m \pm \frac{\Delta_{\psi_m}}{2} \quad (101)$$

and where "+" is used for the top edge and "–" is used for the bottom edge. Finally, for the grid signals at the corners

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\psi_m}} = \left(\frac{\beta_{1m}}{2d_{\psi_m}} - \frac{\beta_{0m}\beta_{3m}}{d_{\phi_m}}\left(\frac{\Delta_{\phi_m}}{d_{\phi_m}} + 1\right)\right)\mathbf{e}_{mi}\mathbf{e}_{mi}^H \pm \frac{\beta_{3m}}{2}(\mathbf{B}_{mi}\mathbf{e}_{mi}\mathbf{e}_{mi}^H - \mathbf{e}_{mi}\mathbf{e}_{mi}^H\mathbf{B}_{mi}). \quad (102)$$

Having differentiated the covariance matrix in terms of all the component signals of the grid, the gradient $G(\Delta_{\phi_m})$ is computed using (81) where $\partial \mathbf{C}_{mi}/\partial \Delta_{\phi_m}$ is defined by one of equations (92), (93), (96), or (97) depending on the location of the signal in the grid.

Similarly, the gradient $G(\Delta_{\psi_m})$ is also computed using (81) where $\partial \mathbf{C}_{mi}/\partial \Delta_{\psi_m}$ is computed using one of equations (98), (99), (100), or (102) depending on the location of the signal in the grid.

4.4.2 The Raised-Cosine Elliptical Grid

Figure 3 shows both the general shape of an elliptical grid using a raised-cosine taper function and the arrangement of point-source signals in one row of the grid (the arrangement in a column is the same). With the taper function going to zero at the edges of the region, no special edge spacings are required to introduce or remove new signals. However, for smaller spread values it becomes necessary to reduce the grid spacings so that at least three point-source signals are used in the center row or center column of the grid. Otherwise the resultant signal could end up being composed of a signal point-source signal in each row and/or column with no way to compute the spread gradient. Additionally, for signals where the spread region drops below the horizon, or above vertical, modifications are required which are discussed in Section 4.5.

For convenience, it will be initially assumed that both the azimuth and elevation spread values (Δ_ϕ and Δ_ψ) are large enough that the spacing parameters d_ϕ and d_ψ can be chosen to be some constant fraction of the horizontal and vertical beamwidths of the antenna array. The case involving smaller spread values will then be dealt with afterwards.

The taper function for the elliptical grid is defined by

$$\beta_{mi} = \begin{cases} c_m(1+\cos(2\pi r_{mi})) & \text{if } r_{mi} \leq 0.5 \\ \\ 0 & \text{otherwise} \end{cases} \tag{103}$$

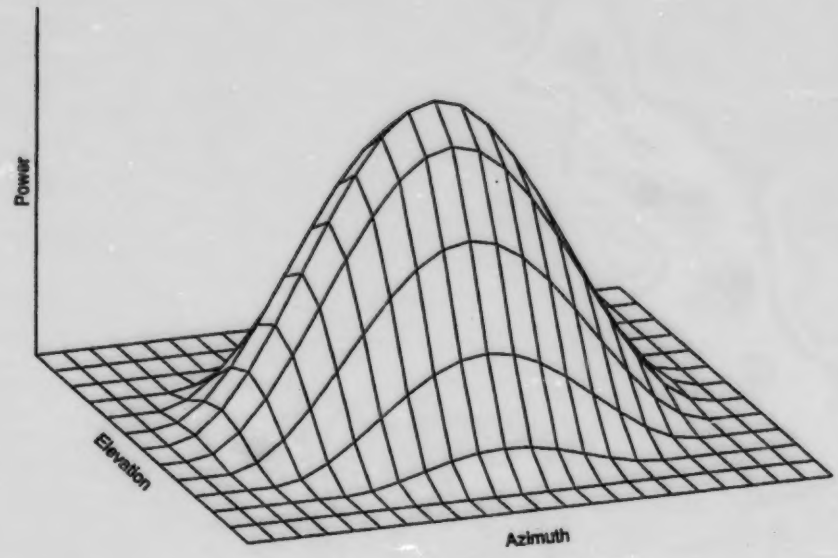
where β_{mi} is used as a shorthand way of writing $\beta_m(\Phi_{mi}, \Psi_{mi})$, c_m is a normalization factor so that $\sum_{i=1}^{M_g} \beta_{mi} = 1$ and can be determined from

$$c_m = \frac{1}{\sum_{k=1}^{M_g}(1+\cos(2\pi r_{mk}))} \tag{104}$$

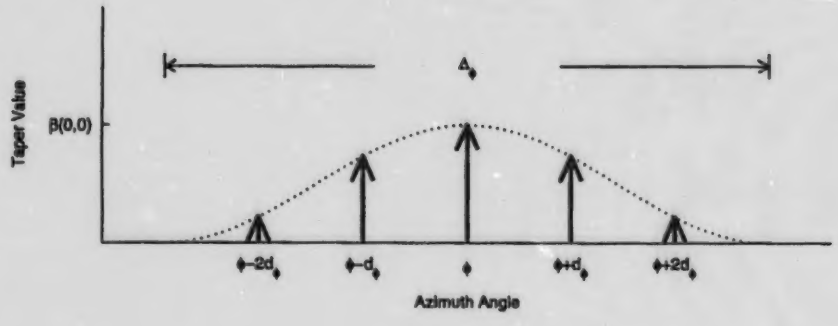
and r_{mi} is the normalized distance from the center of the grid expressed mathematically as

$$r_{mi} = \sqrt{\Phi_{mi}^2 + \Psi_{mi}^2} = \sqrt{\left(\frac{\phi_m - \phi_{mi}}{\Delta_{\phi_m}}\right)^2 + \left(\frac{\psi_m - \psi_{mi}}{\Delta_{\psi_m}}\right)^2}. \tag{105}$$

To avoid unnecessary computations, the grid is only generated in the elliptical region defined by $r_{mi} \leq \frac{1}{2}$. The resultant elliptical grid then measures Δ_{ϕ_m} along the azimuth axis from left to right and Δ_{ψ_m} along the elevation axis from top to bottom (or slightly less than these dimensions depending on the spacing parameters). Additionally, if the



(a)



(b)

Figure 3: Geometry of the raised-cosine elliptical grid showing (a) the general shape, and (b) an example of the position and power of the point-source signals making up one row of a grid.

grid extends below 0° or above 90° in elevation (which is an unrealistic condition for the kinds of signals being dealt with in this report), these parts of the grid are removed.

Based on the fixed spacing for all the signals used in the grid, the intermediate result (91) found for the rectangular grid is also appropriate for the elliptical grid and, after making the appropriate modifications, is shown here as

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_m} = \frac{\partial \beta_{mi}}{\partial \Delta_m} \mathbf{e}_{mi} \mathbf{e}_{mi}^H. \qquad (106)$$

Using the definition for β_{mi} for the elliptical grid, and letting $\Delta_m = \Delta_{\phi_m}$ then

$$\frac{\partial \beta_{mi}}{\partial \Delta_m} = \frac{4\pi^2 c_m}{\Delta_{\phi_m}} \left(\Phi_{mi}^2 \operatorname{sinc}(2r_{mi}) - \beta_{mi} \sum_{k=1}^{M_g} \Phi_{mk}^2 \operatorname{sinc}(2r_{mk}) \right) \qquad (107)$$

and the azimuth spread gradient (106) becomes

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\phi_m}} = \frac{4\pi^2 c_m}{\Delta_{\phi_m}} \left(\Phi_{mi}^2 \operatorname{sinc}(2r_{mi}) - \beta_{mi} \sum_{k=1}^{M_g} \Phi_{mk}^2 \operatorname{sinc}(2r_{mk}) \right) \mathbf{e}_{mi} \mathbf{e}_{mi}^H \qquad (108)$$

where $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$ and $\operatorname{sinc}(0) = 1$. Letting $\Delta_m = \Delta_{\psi_m}$ the equivalent elevation spread gradient becomes

$$\frac{\partial \mathbf{C}_{mi}}{\partial \Delta_{\psi_m}} = \frac{4\pi^2 c_m}{\Delta_{\psi_m}} \left(\Psi_{mi}^2 \operatorname{sinc}(2r_{mi}) - \beta_{mi} \sum_{k=1}^{M_g} \Psi_{mk}^2 \operatorname{sinc}(2r_{mk}) \right) \mathbf{e}_{mi} \mathbf{e}_{mi}^H. \qquad (109)$$

These last two expressions can then be substituted back into (81) to compute the gradients $G(\Delta_{\phi_m})$ and $G(\Delta_{\psi_m})$, respectively.

If the azimuth and/or elevation spread value is reduced sufficiently, the above approach fails because the number of rows and/or columns in the grid is reduced to one. For example, if the spread $\Delta_\phi < 2d_\phi$ in Figure 3, then there would be a single point-source signal at ϕ in each row of the grid. At this point, then, the corresponding value of β_{mi} will not change as the given spread is further reduced so that the corresponding differential becomes zero. This condition occurs when

$$\Delta_{\phi_m} < 2d_{\phi_m} \qquad (110)$$

and/or

$$\Delta_{\psi_m} < 2d_{\psi_m} \qquad (111)$$

where d_{ϕ_m} and d_{ψ_m} are fixed values.

To overcome this difficulty, modeling of the spread signal is changed so that the central row and column contain at least three point-source signals. This is done by using the adjustable azimuth grid spacing

$$d'_{\phi_m} = \frac{1}{4}\Delta_{\phi_m} \qquad \text{for } \Delta_{\phi_m} < 4d_{\phi_m} \tag{112}$$

instead of the fixed spacing d_{ϕ_m}, and using the adjustable elevation grid spacing

$$d'_{\psi_m} = \frac{1}{4}\Delta_{\psi_m} \qquad \text{for } \Delta_{\psi_m} < 4d_{\psi_m} \tag{113}$$

instead of the fixed spacing d_{ψ_m}. An example for the case of narrow spreading in azimuth is shown in Figure 4. The conditions placed on d'_{ϕ_m} and d'_{ψ_m} in the preceding expressions have been chosen to ensure continuity between the narrow spread model and the wide spread model since for $\Delta_{\phi_m} = 4d_{\phi_m}$ and/or $\Delta_{\psi_m} = 4d_{\psi_m}$, the two models are identical.

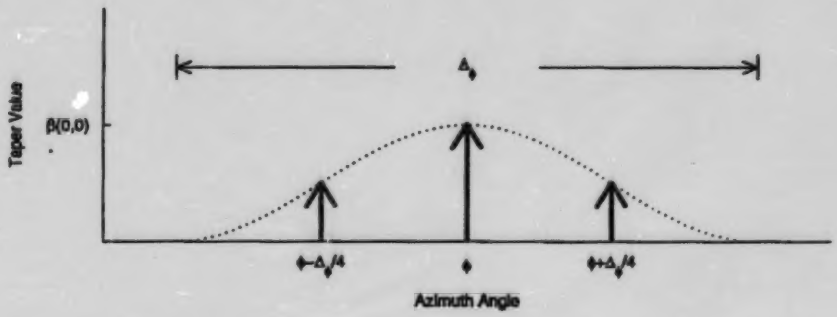


Figure 4: Position of point-source signals in a raised-cosine elliptical grid used for narrow spread signals.

For the narrow spread model, the angles of the outer point-source signals are always in the same relative position with respect to the central point-source and the spread angle, so that the corresponding values of β_{mi} become constant. For example, assuming that $\Delta_{\phi_m} < 4d_\phi$ then the values of ϕ_{mi} are chosen from the set

$$\{\phi_m - \frac{\Delta_{\phi_m}}{4},\ \phi_m,\ \phi_m + \frac{\Delta_{\phi_m}}{4}\}. \tag{114}$$

Since $\Phi_{mi} = (\phi_m - \phi_{mi})/\Delta_{\phi_m}$, then the corresponding values of Φ_{mi} are chosen from the

set

$$\{-\frac{1}{4},\ 0,\ \frac{1}{4}\} \tag{115}$$

which is independent of Δ_{ϕ_m}. Based on the relationship between Φ_{mi} and β_{mi} expressed in (103)-(105), then β_{mi} will also be independent of the azimuth spread. Similarly, assuming that $\Delta_{\phi_m} < 4d_\phi$ then the values of ϕ_{mi} are chosen from the set

$$\{\psi_m - \frac{\Delta_{\psi_m}}{4},\ 0,\ \psi_m - \frac{\Delta_{\psi_m}}{4}\}. \tag{116}$$

Since $\Psi_{mi} = (\psi_m - \psi_{mi})/\Delta_{\psi_m}$, the resultant values of Ψ_{mi} are also given by (115), and as before, this leads to β_{mi} being independent of the elevation spread.

Deriving the gradient for the narrow spread case is very similar to deriving the bearing gradient since only the point-source signal directions are affected by changes in spread, and not the value of the taper function β_{mi}. Hence using the results from the bearing gradient case, then for $\Delta_{\phi_m} < 4d_\phi$ the azimuth spread gradient is then given by

$$\frac{\partial \mathbf{C}}{\partial \Delta_{\phi_m}} = s_m \sum_{i=1}^{M_g} \beta_{mi} \left(\frac{\partial \mathbf{e}_{mi}}{\partial \Delta_{\phi_m}} \mathbf{e}_{mi}^H + \mathbf{e}_{mi} \frac{\partial \mathbf{e}_{mi}^H}{\partial \Delta_{\phi_m}}\right) \tag{117}$$

where

$$\frac{\partial \mathbf{e}_{mi}}{\partial \Delta_{\phi_m}} = \begin{cases} -\frac{1}{4}\mathbf{A}_{mi}\mathbf{e}_{mi} & \text{for } \Phi_{mi} = -\frac{1}{4} \\ 0 & \text{for } \Phi_{mi} = 0 \\ \frac{1}{4}\mathbf{A}_{mi}\mathbf{e}_{mi} & \text{for } \Phi_{mi} = \frac{1}{4} \end{cases} \tag{118}$$

For $\Delta_{\psi_m} < 4d_\psi$ the elevation spread gradient is given by

$$\frac{\partial \mathbf{C}}{\partial \Delta_{\psi_m}} = s_m \sum_{i=1}^{M_g} \beta_{mi} \left(\frac{\partial \mathbf{e}_{mi}}{\partial \Delta_{\psi_m}} \mathbf{e}_{mi}^H + \mathbf{e}_{mi} \frac{\partial \mathbf{e}_{mi}^H}{\partial \Delta_{\psi_m}}\right) \tag{119}$$

where

$$\frac{\partial \mathbf{e}_{mi}}{\partial \Delta_{\psi_m}} = \begin{cases} -\frac{1}{4}\mathbf{B}_{mi}\mathbf{e}_{mi} & \text{for } \Psi_{mi} = -\frac{1}{4} \\ 0 & \text{for } \Psi_{mi} = 0 \\ \frac{1}{4}\mathbf{B}_{mi}\mathbf{e}_{mi} & \text{for } \Psi_{mi} = \frac{1}{4} \end{cases} \tag{120}$$

The matrices $\mathbf{A}_{mi}$ and $\mathbf{B}_{mi}$ were defined previously by (73) and (76), respectively, in

4.5 Out-of-Bounds Signals

This section deals with the problem of modeling the spread region with a planar array, when part of the grid is adjusted so that it falls below the horizon or above vertical (out-of-bounds). Without compensation, the out-of-bound signals are folded back in-bounds corrupting the shape of the signal model spread region (i.e., grid signals below the horizon act as if the elevation angle is $-\psi_{mi}$, while signals above the vertical act as if the elevation angle is $\psi_{mi} - 90°$. The compensation required depends on the taper function used.

4.5.1 The Uniform Rectangular Grid

For the uniform rectangular taper function, the problem of out-of-bound grid signals can be easily handled by readjusting the elevation angle and spread parameters. For spreading below the horizon ($\psi_m - \Delta_{\psi_m}/2 < 0°$) the modifications are

$$\psi'_m = \frac{1}{2}(\psi_m + \frac{\Delta_{\psi_m}}{2}) \tag{121}$$

$$\Delta'_{\psi_m} = 2\psi'_m \tag{122}$$

where the ψ'_m and Δ'_{ψ_m} are the updated values, and for spreading above vertical ($\psi_m - \Delta_{\psi_m}/2 > 90°$) the modifications are

$$\psi'_m = \frac{1}{2}(\psi_m - \frac{\Delta_{\psi_m}}{2}) + 45° \tag{123}$$

$$\Delta'_{\psi_m} = 2(90° - \psi'_m). \tag{124}$$

These readjustments avoid the need to derive expression (77) since the readjustment can be done without affecting the shape or distribution of the spread region.

4.5.2 The Raised-Cosine Elliptical Grid

For the raised-cosine elliptical taper function things are not as easily handled as for the uniform rectangular taper function due to its nonuniform nature. Instead the part of the grid that is out-of-bounds is first removed. However, this leads to discontinuities in the gradient functions when grid signals are added or deleted in this manner. To handle this transition, a new row of grid signals are introduced right at the boundary as shown in Figure 5. The power of these boundary signals are determined in much the same way as done for the edge signals of the uniform rectangular grid. The model covariance matrix is then formed using

$$\mathbf{C} = \sigma^2 \mathbf{C}_\eta + \sum_{m=1}^{M} s_m \left(\sum_j \beta_{mj} \mathbf{e}_{mj} \mathbf{e}_{mj}^H + \sum_k \beta_{mk} \mathbf{e}_{mk} \mathbf{e}_{mk}^H \right) \tag{125}$$

where the shorthand form β_{mi} is again used for $\beta_m(\Phi_{mi}, \Psi_{mj})$, and the summation with respect to the index j is assumed to include all boundary grid signals while the summation with respect to the index k is assumed to include all other grid signals which are within the boundaries. To remain consistent throughout the rest of this section, the indices j and j' will be used to denote a parameter associated with the grid signals on the boundary, and k and k' will be used to denote a parameter associated with grid signals inside the boundaries. It is also assumed that the elevation spread region is narrow enough that the removed grid rows are either all below the horizon or all above vertical, but not both.

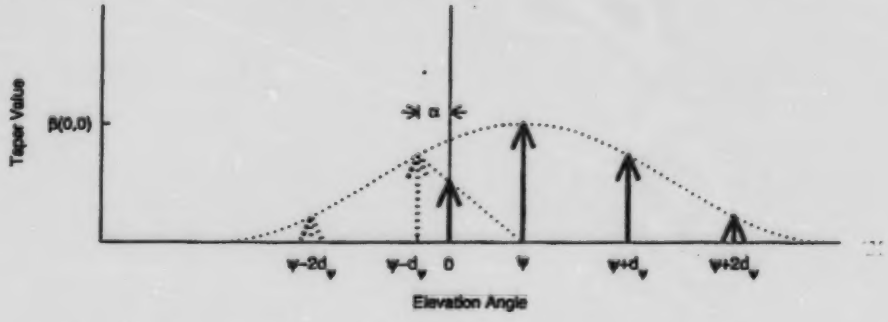


Figure 5: Position of point-source signals in a raised-cosine elliptical grid where the spread region goes below the horizon. Dotted arrows show angular positions of rows that have been removed.

The value of the taper function for β_{mk} is computed in the same way as before, namely,

$$\beta_{mk} = c_m(1 + \cos 2\pi r_{mk}) \tag{126}$$

while the adjusted value of the taper function β_{mj} is computed using

$$\beta_{mj} = c_m(1 + \frac{\alpha_m}{d_m})(1 + \cos 2\pi\rho_{mj}) \tag{127}$$

where d_m is the grid elevation spacing parameter defined as

$$d_m = \begin{cases} d_{\psi_m} & \text{for } \Delta_{\psi_m} \geq 4d_{\psi_m} \\ \\ \frac{1}{4}\Delta_{\psi_m} & \text{for } \Delta_{\psi_m} < 4d_{\psi_m} \end{cases} \tag{128}$$

and α_m is the angular elevation difference between the boundary and the first deleted grid row as shown in Figure 5. More specifically, it is defined according to

$$\alpha_m = \begin{cases} \psi_{mv} & \text{below the horizon} \\ \\ 90° - \psi_{mv} & \text{above vertical} \end{cases} \tag{129}$$

where ψ_{mv} is defined as the elevation angle of the removed (or virtual) row of grid signals closest too, but below the horizon or above vertical (the corresponding elevation angles ψ_{mj} are either 0° or 90°). The normalizing constant then becomes

$$c_m = \frac{1}{\sum_k (1 + \cos 2\pi\tau_{mk}) + (1 + \frac{\alpha_m}{d_m}) \sum_{j'} (1 + \cos 2\pi\rho_{mj'})}. \tag{130}$$

The definition for τ_{mk} is given by (105) and the definition of ρ_{mj} is the same as τ_{mk} except that the elevation angle used in this case is ψ_{mv} (instead of ψ_{mj}), so the definition becomes

$$\rho_{mj} = \sqrt{\Phi_{mj}^2 + \Upsilon_{mj}^2} = \sqrt{\left(\frac{\phi_m - \phi_{mj}}{\Delta_{\phi_m}}\right)^2 + \left(\frac{\psi_m - \psi_{mv}}{\Delta_{\psi_m}}\right)^2}. \tag{131}$$

Additionally, Υ_{mj} is used in place of Ψ_{mj} in the above expression to indicate the modification to the elevation angle.

Using these new definitions, the calculation of the noise, signal power, azimuth bearing, and azimuth spread gradients proceed in the same manner as before. The gradients for the elevation bearing and elevation spread, however, need modification.

Starting with the elevation bearing, and given the modifications made to the signal

grid, the gradient given by (77) can be expressed as

$$G(\psi_m) = s_m \text{trace}\left(\mathbf{P} \sum_j \frac{\partial \beta_{mj}}{\partial \psi_m} \mathbf{e}_{mj} \mathbf{e}_{mj}^H + \mathbf{P} \sum_k \left[\frac{\partial \beta_{mk}}{\partial \psi_m} \mathbf{e}_{mk} \mathbf{e}_{mk}^H + \beta_{mk} (\mathbf{B}_{mk} \mathbf{e}_{mk} \mathbf{e}_{mk}^H - \mathbf{e}_{mk} \mathbf{e}_{mk}^H \mathbf{B}_{mk}) \right] \right). \quad (132)$$

Performing the derivative operations indicated then

$$\begin{aligned} \frac{\partial \beta_{mj}}{\partial \psi_m} &= \frac{\pm c_m}{d_m}(1 + \cos 2\pi \rho_{mj}) + \frac{\partial c_m}{\partial \psi_m}(1 + \cos 2\pi \rho_{mj})(1 + \frac{\alpha_m}{d_m}) \\ &= \frac{\pm c_m}{d_m}(1 + \cos 2\pi \rho_{mj}) - \frac{\pm c_m^2}{d_m}(1 + \cos 2\pi \rho_{mj})(1 + \frac{\alpha_m}{d_m}) \sum_{j'} (1 + \cos 2\pi \rho_{mj'}) \\ &= \frac{\pm \beta_{mj}}{d_m + \alpha_m}(1 - \sum_{j'} \beta_{mj'}) \end{aligned} \quad (133)$$

$$\begin{aligned} \frac{\partial \beta_{mk}}{\partial \psi_m} &= \frac{\partial c_m}{\partial \psi_m}(1 + \cos 2\pi r_{mk}) \\ &= (-1)\frac{\pm c_m^2}{d_m}(1 + \cos 2\pi r_{mk}) \sum_{j'} (1 + \cos 2\pi r_{mj'}) \\ &= (-1)\frac{\pm \beta_{mk}}{d_m + \alpha_m} \sum_{j'} \beta_{mj'} \end{aligned} \quad (134)$$

where given a choice, "+" is used for below horizon conditions and "−" is used for above vertical conditions. The elevation gradient expressed in (132) can now be computed by substituting (133) and (134) for the derivatives.

Before continuing on to the spread gradients, it is useful to distinguish between the two separate cases, namely modeling signals with a wide elevation spread where

$$\Delta_{\psi_m} \geq 4 d_{\psi_m}, \quad (135)$$

and modeling signals with a narrow elevation spread where

$$\Delta_{\psi_m} < 4 d_{\psi_m}. \quad (136)$$

The main difference is that in the wide case the grid elevation spacing is fixed while in the narrow case the grid elevation spacing changes as a function of the elevation spread angle (see (128)). Since these differences lead to different derivations, the two cases are handled separately.

Starting with the wide elevation spread case, the gradient is given by

$$G(\Delta_{\psi_m}) = s_m \text{trace}\left(\mathbf{P}\sum_j \frac{\partial \beta_{mj}}{\partial \Delta_{\psi_m}} \mathbf{e}_{mj}\mathbf{e}_{mj}^H + \mathbf{P}\sum_k \frac{\partial \beta_{mk}}{\partial \Delta_{\psi_m}} \mathbf{e}_{mk}\mathbf{e}_{mk}^H\right). \quad (137)$$

The derivatives of β_{mj} and β_{mk} are practically identical to the corresponding derivative for β_{mi} in the previous section (107) except for minor modifications to account for the boundary signals. With these modifications, then

$$\frac{\partial \beta_{mj}}{\partial \Delta_{\psi_m}} = \frac{4\pi^2 c_m}{\Delta_{\psi_m}}\left(\Upsilon_{mj}^2(1+\frac{\alpha_m}{d_{\psi_m}})\,\text{sinc}\,2\rho_{mj} - \beta_{mj}(1+\frac{\alpha_m}{d_{\psi_m}})\sum_{j'}\Upsilon_{mj'}^2\,\text{sinc}\,2\rho_{mj'} - \beta_{mj}\sum_k \Psi_{mk}^2\,\text{sinc}\,2r_{mk}\right) \quad (138)$$

$$\frac{\partial \beta_{mk}}{\partial \Delta_{\psi_m}} = \frac{4\pi^2 c_m}{\Delta_{\psi_m}}\left(\Psi_{mk}^2\,\text{sinc}\,2r_{mk} - \beta_{mk}(1+\frac{\alpha_m}{d_{\psi_m}})\sum_{j'}\Upsilon_{mj'}^2\,\text{sinc}\,2r_{mj'} - \beta_{mk}\sum_{k'}\Psi_{mk'}^2\,\text{sinc}\,2r_{mk'}\right). \quad (139)$$

The gradient for the azimuth spread (wide case) can then be computed by substituting (138) and (139) for the appropriate derivatives in (137).

For the narrow azimuth spread case the gradient is given by

$$G(\Delta_{\psi_m}) = s_m \text{trace}\left(\mathbf{P}\sum_j \frac{\partial \beta_{mj}}{\partial \Delta_{\psi_m}} \mathbf{e}_{mj}\mathbf{e}_{mj}^H + \mathbf{P}\sum_k \left[\frac{\partial \beta_{mk}}{\partial \Delta_{\psi_m}}\mathbf{e}_{mk}\mathbf{e}_{mk}^H + \beta_{mk}(\frac{\partial \mathbf{e}_{mk}}{\partial \Delta_{\psi_m}}\mathbf{e}_{mk}^H + \mathbf{e}_{mk}\frac{\partial \mathbf{e}_{mk}^H}{\partial \Delta_{\psi_m}})\right]\right) \quad (140)$$

where $\partial \mathbf{e}_{mk}/\partial \Delta_{\psi_m}$ is defined by (120). Using the fact that $\psi_{mj} = \psi_m - d'_{\psi_m}$ for below

horizon signals, $\psi_{mj} = \psi_m + d'_{\psi_m}$ for above vertical signals, and $d'_{\psi_m} = \Delta_{\psi_m}/4$ then

$$\beta_{mj} = c_m \frac{4\nu_m}{\Delta_{\psi_m}}(1 + \cos 2\pi\tau_{mj}) \tag{141}$$

and

$$c_m = \frac{1}{\sum_k (1 + \cos 2\pi\tau_{mk}) + \frac{4\nu_m}{\Delta_{\psi_m}} \sum_{j'} (1 + \cos 2\pi\tau_{mj'})} \tag{142}$$

where

$$\nu_m = \begin{cases} \psi_m & \text{below the horizon} \\ 90^\circ - \psi_m & \text{above vertical} \end{cases} \tag{143}$$

The differentiation of β_{mj} then proceeds as

$$\begin{aligned} \frac{\partial \beta_{mj}}{\partial \Delta_{\psi_m}} &= \frac{\partial c_m}{\partial \Delta_{\psi_m}} \frac{4\nu_m}{\Delta_{\psi_m}}(1 + \cos 2\pi\tau_{mj}) - c_m \frac{4\nu_m}{\Delta^2_{\psi_m}}(1 + \cos 2\pi\tau_{mj}) \\ &= c_m^2 \frac{16\nu_m^2}{\Delta^3_{\nu_m}}(1 + \cos 2\pi\tau_{mj}) \sum_{j'} (1 + \cos 2\pi\tau_{mj'}) - \frac{\beta_{mj}}{\Delta_{\psi_m}} \\ &= \frac{\beta_{mj}}{\Delta_{\psi_m}} \sum_{j'} \beta_{mj'} - \frac{\beta_{mj}}{\Delta_{\psi_m}}. \end{aligned} \tag{144}$$

In a similar manner, the differentiation of β_{mk} yields

$$\frac{\partial \beta_{mk}}{\partial \Delta_{\psi_m}} = \frac{\beta_{mk}}{\Delta_{\psi_m}} \sum_{j'} \beta_{mj'}. \tag{145}$$

Based on the derivations here and those in the previous section, the gradient for the azimuth spread (narrow case) can be computed from (140) by substituting (120), (144), and (145) as appropriate.

5.0 SML ALGORITHM SUMMARY

The following two sections summarize the main SML routine and the gradient ascent routine. The gradient ascent routine was summarized separately owing to its complexity, otherwise the resultant combined routine would have been more difficult to follow.

In the main SML routine, two different constant values are introduced which affect algorithm performance. The first is a 30% threshold level used to determine whether a maximum value found for the function $S_m(\phi_m, \psi_m)$ is significant with respect to the global maximum value. A higher value could be used to reduce the number of maximum values which exceed this threshold, however through numerous simulations it has been found that this reduces accuracy for conditions where angular separations between signals are a beamwidth or less.

The second constant value is the grid spacing which is chosen as 30% of the array beamwidth. The reasons for this choice are discussed later in Section 6.1.

5.1 Main SML Routine

1. Compute the data covariance matrix using (11)

$$\mathbf{R} = \frac{1}{K}\mathbf{X}\mathbf{X}^H. \tag{146}$$

2. Set the stage counter index: $m = 0$.

3. Estimate the normalized noise covariance matrix based on using theoretical considerations or practical measurements. For example, the theoretical covariance for white Gaussian noise is found according to (18)

$$\mathbf{C}_\eta = \frac{1}{N}\mathbf{I}_N. \tag{147}$$

Alternatively, if measurements of noise only data are available then the normalized noise covariance matrix is given by (19)

$$\mathbf{C}_\eta = \frac{\mathbf{X}\mathbf{X}^H}{\text{trace}(\mathbf{X}\mathbf{X}^H)}. \tag{148}$$

4. For the moment, assume the data contains no signals and estimate the noise power using (35)

$$\sigma^2 = \frac{1}{N}\text{trace}(\mathbf{R}\mathbf{C}_\eta^{-1}), \tag{149}$$

then generate the initial model covariance matrix using (32) with $\mathbf{C} = \mathbf{C}_0$ giving

$$\mathbf{C}_0 = \sigma^2 \mathbf{C}_\eta. \tag{150}$$

5. Increment the index: $m = m + 1$.

6. Create a list of candidate azimuth and elevation bearings (within the angular ranges of interest) yielding local maximum values within 30% of the global maximum of the function defined by (48) where $\mathbf{C}_\alpha = \mathbf{C}_{m-1}$, or

$$S_m(\phi_m, \psi_m) = \frac{\mathbf{e}_m^H \mathbf{C}_{m-1}^{-1} \mathbf{R} \mathbf{C}_{m-1}^{-1} \mathbf{e}_m}{\mathbf{e}_m^H \mathbf{C}_{m-1}^{-1} \mathbf{e}_m}. \tag{151}$$

The matrix $\mathbf{C}_{m-1}$ represents the model covariance matrix generated at stage $m-1$, and the array response or steering vector is given by (15)

$$\mathbf{e}_m = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j\frac{2\pi}{\lambda}(x_0 \sin\phi_m \cos\psi_m + y_0 \cos\phi_m \cos\psi_m)} \\ e^{j\frac{2\pi}{\lambda}(x_1 \sin\phi_m \cos\psi_m + y_1 \cos\phi_m \cos\psi_m)} \\ \vdots \\ e^{j\frac{2\pi}{\lambda}(x_{N-1} \sin\phi_m \cos\psi_m + y_{N-1} \cos\phi_m \cos\psi_m)} \end{bmatrix}. \tag{152}$$

7. For each list entry containing candidate signal bearings, include the corresponding signal power calculated using(40), (44), and (47), or equivalently

$$s_m = \frac{S_m(\phi_m, \psi_m) - 1}{\mathbf{e}_m^H \mathbf{C}_{m-1}^{-1} \mathbf{e}_m}, \tag{153}$$

and the spread values

$$\Delta_{\phi_m} = \Delta_{\psi_m} = 0. \tag{154}$$

Also include all the signal and noise model parameters from the optimum model computed at stage $m-1$. Hence each list entry will contain the initial estimates for a possible solution involving one new signal, $m-1$ previously established signals, and noise.

8. If for any list entry the new candidate signal bearing coincides (collides) with the

scattering region of any of the established signals, perform the following:

(a) create a new list entry by repeating all the parameters from the colliding entry;

(b) reset the signal power of all colliding signals in the new entry to the average signal power of these same signals; and

(c) set the azimuth and elevation spread values of all colliding signals to zero.

9. Refine the noise and signal parameter values of each list entry using the gradient ascent technique. If the uniform rectangular taper function is employed, make adjustments to the appropriate parameters if out-of-bound conditions are detected as outlined in Section 4.5.1.

10. Update the model covariance matrix $\mathbf{C}_m$ associated with each list entry according to the current model parameter refinements using either (31)

$$\mathbf{C}_m = \sigma^2 \mathbf{C}_\eta + \sum_{n=1}^{m} s_n \sum_{i=1}^{M_g} \beta_n(\Phi_{ni}, \Psi_{ni}) \mathbf{e}_{ni} \mathbf{e}_{ni}^H \tag{155}$$

where the values of $\beta_n(\Phi_{ni}, \Psi_{ni})$ are computed as discussed in Section 4.4, or if the raised-cosine elliptical taper function is employed and out-of-bounds conditions are detected then use (125)

$$\mathbf{C}_m = \sigma^2 \mathbf{C}_\eta + \sum_{n=1}^{m} s_n \left(\sum_j \beta_{nj} \mathbf{e}_{nj} \mathbf{e}_{nj}^H + \sum_k \beta_{nk} \mathbf{e}_{nk} \mathbf{e}_{nk}^H \right) \tag{156}$$

where β_{nj} and β_{nk} are computed as discussed in Section 4.5.2.

11. Also compute the cost function value of the refined model for each list entry using (10) with $\mathbf{C} = \mathbf{C}_m$ which yields

$$L_m = -\ln(\det \mathbf{C}_m) - \text{trace}(\mathbf{R}\mathbf{C}_m^{-1}). \tag{157}$$

Note that if the number of list entries is excessive, pruning of this list to improve computational speed can be accomplished by performing Steps 9-11 using only 20 iterations of the gradient ascent technique and then retaining the list entries corresponding to the four highest cost values. Steps 9-11 are then repeated for the four remaining entries using the full number of iterations for the gradient ascent technique.

12. Select the model parameters in the list corresponding to the highest cost function value. These are considered the optimum model parameters estimates for stage m.

13. Repeat Steps 5-13 as long as $m \leq M$.

14. Output the model parameter results from stage M.

5.2 Gradient Ascent Routine

1. Save the value of the cost function for the current unrefined model: $L_{max} = L_m$.

2. Initialize the gradient ascent loop parameters according to the following

$$\begin{aligned} loop &= 0 \\ \mu_\eta &= \xi \\ \mu_{s_k} &= \xi \\ \mu_{\phi_k} &= \xi \\ \mu_{\psi_k} &= \xi \\ \mu_{\Delta_{\phi_k}} &= \xi \\ \mu_{\Delta_{\psi_k}} &= \xi \end{aligned}$$

where $k = 1, ..., m$ and ξ is the initial step size. A suggested approach to determine an appropriate initial step size is to compute the initial gradients of the azimuth and elevation bearings and the initial gradients of the bearing spreads for all signals in the current model using Steps 4-5. The maximum absolute value g_{max} is selected from these gradients and the initial step size is then computed using $\xi = 0.05\phi_{bw}/g_{max}$ where ϕ_{bw} is the 3 dB azimuth beamwidth of the antenna array. This restricts the initial adjustment of the bearing and spread parameters to no more than 5% of the azimuth beamwidth.

3. Increment the loop counter: $loop = loop + 1$.

4. Compute the quantity (64)

$$\mathbf{P} = (\mathbf{C}_m^{-1}\mathbf{R} - \mathbf{I})\mathbf{C}_m^{-1}. \tag{158}$$

5. Compute the gradients

$$G(\sigma^2) = \text{trace}(\mathbf{P}\mathbf{C}_\eta) \tag{159}$$

$$G(s_k) = \text{trace}(\mathbf{P}\sum_{i=1}^{M_g}\beta_k(\Phi_{ki},\Psi_{ki})\mathbf{e}_{ki}\mathbf{e}_{ki}^H) \tag{160}$$

$$G(\phi_k) = s_k\text{trace}(\mathbf{P}\sum_{i=1}^{M_g}\beta_k(\Phi_{ki},\Psi_{ki})(\mathbf{A}_{ki}\mathbf{e}_{ki}\mathbf{e}_{ki}^H - \mathbf{e}_{ki}\mathbf{e}_{ki}^H\mathbf{A}_{ki})) \tag{161}$$

$$G(\psi_k) = s_k\text{trace}(\mathbf{P}\sum_{i=1}^{M_g}\beta_k(\Phi_{ki},\Psi_{ki})(\mathbf{B}_{ki}\mathbf{e}_{ki}\mathbf{e}_{ki}^H - \mathbf{e}_{ki}\mathbf{e}_{ki}^H\mathbf{B}_{ki})) \tag{162}$$

(from (68), (70), (74), and (75), respectively) for $k = 1, ..., m$ and where $\mathbf{A}_{ki}$ and $\mathbf{B}_{ki}$ are diagonal matrices defined by (73) and (76) respectively, which are repeated here as

$$\text{diag}(\mathbf{A}_{ki}) = j\frac{2\pi}{\lambda}\begin{bmatrix} x_0\cos\phi_{ki}\cos\psi_{ki} - y_0\sin\phi_{ki}\cos\psi_{ki} \\ x_1\cos\phi_{ki}\cos\psi_{ki} - y_1\sin\phi_{ki}\cos\psi_{ki} \\ \vdots \\ x_{N-1}\cos\phi_{ki}\cos\psi_{ki} - y_{N-1}\sin\phi_{ki}\cos\psi_{ki} \end{bmatrix} \tag{163}$$

$$\text{diag}(\mathbf{B}_{ki}) = -j\frac{2\pi}{\lambda}\begin{bmatrix} x_0\sin\phi_{ki}\sin\psi_{ki} + y_0\cos\phi_{ki}\sin\psi_{ki} \\ x_1\sin\phi_{ki}\sin\psi_{ki} + y_1\cos\phi_{ki}\sin\psi_{ki} \\ \vdots \\ x_{N-1}\sin\phi_{ki}\sin\psi_{ki} + y_{N-1}\cos\phi_{ki}\sin\psi_{ki} \end{bmatrix}. \tag{164}$$

Since the gradients for $G(\Delta_{\phi_k})$ and $G(\Delta_{\psi_k})$ cannot be easily summarized in one line, refer to Section 4.4. Additionally, for out-of-bounds conditions (which affects $G(\psi_k)$ and $G(\Delta_{\psi_k})$), refer to Section 4.5.

6. Update the model parameters by adapting (54) as appropriate and letting $i = loop$ to get

$$\sigma^2 \rightarrow \sigma^2 + \mu_\eta G(\sigma^2) \tag{165}$$

$$s_k \rightarrow s_k + \mu_{s_k} G(s_k) \tag{166}$$

$$\phi_k \rightarrow \phi_k + \mu_{\phi_k} G(\phi_k) \tag{167}$$

$$\psi_k \rightarrow \psi_k + \mu_{\psi_k} G(\psi_k) \tag{168}$$

$$\Delta_{\phi_k} \rightarrow \Delta_{\phi_k} + \mu_{\Delta_{\phi_k}} G(\Delta_{\phi_k}) \tag{169}$$

$$\Delta_{\psi_k} \rightarrow \Delta_{\psi_k} + \mu_{\Delta_{\psi_k}} G(\Delta_{\psi_k}) \tag{170}$$

for $k = 1, ..., m$. If the uniform-rectangular taper function is used and out-of-bounds conditions are detected, adjust the appropriate model parameters as outlined in Section 4.5.1.

7. Compute the model covariance matrix using (31)

$$\mathbf{C}_m = \sigma^2 \mathbf{C}_\eta + \sum_{n=1}^{m} s_n \sum_{i=1}^{M_g} \beta_n(\Phi_{ni}, \Psi_{ni}) \mathbf{e}_{ni} \mathbf{e}_{ni}^H \tag{171}$$

where the values of $\beta_n(\Phi_{ni}, \Psi_{ni})$ and the corresponding grid of point-source signals are computed as discussed in Section 4.4 (see equations (82)-(90) for a rectangular grid and equations (103)-(105) for an elliptical grid) Alternatively, if the raised-cosine elliptical taper function is employed and out-of-bounds conditions are detected then using (125)

$$\mathbf{C}_m = \sigma^2 \mathbf{C}_\eta + \sum_{n=1}^{m} s_n \left(\sum_j \beta_{nj} \mathbf{e}_{nj} \mathbf{e}_{nj}^H + \sum_k \beta_{nk} \mathbf{e}_{nk} \mathbf{e}_{nk}^H \right) \tag{172}$$

where β_{nj}, β_{nk} and the corresponding grid of point-source signals are computed as discussed in Section 4.5.2 (see equations (126)-(131)). The grid spacing values are chosen as $d_{\phi_m} = 0.3\phi_{bw}$ and $d_{\psi_m} = 0.3\psi_{bw}$, where ψ_{bw} is the 3 dB elevation beamwidth of the antenna array.

8. Update the cost function using (10)

$$L_m = -\ln(\det \mathbf{C}_m) - \text{trace}(\mathbf{R}\mathbf{C}_m^{-1}). \tag{173}$$

9. If the new cost value is less than the old cost value ($L_m < L_{max}$) then:

   (a) reset the model parameters and the model covariance matrix to their original values at the beginning of the loop;

   (b) reduce all the step size parameters by a factor of 3; and

   (c) if *loop* < 200 go to Step 3, otherwise continue on to Step 10.

Otherwise if ($L_m \geq L_{max}$) then:

(a) save the new cost value: $L_{max} = L_m$;

(b) increase all step size values by a factor of 1.2;

(c) compare all gradient values to the corresponding value from the previous loop and if there has been a sign change (negative to positive or vice versa) decrease the associated step size value by a factor of 3; and

(d) if any of the azimuth or elevation bearings changed more than 5×10^{-6} degrees in Step 8 and $loop < 200$, then go to Step 3, otherwise continue on to Step 10.

10. If $loop < 5$, then reinitialize the step size parameters (using the same method as in Step 2) and go to step 3. Otherwise continue on to the next step.

11. Output the final model parameter results.

6.0 SML CONTROL PARAMETER SETTINGS

In the previous sections, the SML algorithm has been described in some detail. A few of the more practical considerations have not yet been discussed adequately. This includes the control parameter settings for the grid spacing to be used for the signal model, the amount of data that should be collected to form the data covariance matrix, the choice of the taper function, and the assumed number of signals.

To investigate these control parameters, numerical simulations were carried out to determine their most appropriate settings or range of settings. In the simulations, the signals received at the antenna array consisted of a single point-source signal, representing a signal from the desired great circle direction, plus two stronger spread signals, representing signals off the great circle bearing scattered by closely following ionospheric patches. The spread signals were modeled using a large number of point-source signals spread with random amplitudes, phases, and bearings. Unless otherwise specified, the spread signals were set up to have an elliptical shape with a raised-cosine power distribution. The spread signals were also uncorrelated from sample to sample by computing every sample using different random values. Ground reflections and local site multipath effects were not included. Noise was uncorrelated (both spatially and temporally) white Gaussian noise. The relevant parameters, as measured at the receiver, are given in Table 1.

Table 1: Simulation Signal Parameters

Signal	ϕ	ψ	Δ_ϕ	Δ_ψ	Power
1	50°	10°	0°	0°	-20 dB
2	180°	30°	30°	10°	0 dB
3	210°	25°	15°	5°	0 dB
noise	–	–	–	–	-15 dB

The antenna array geometry was as shown in Figure 6. This geometry was investigated in [12] and [13] and found to have very good characteristics for direction finding. Assuming an ideal free space response, for a given signal bearing the 3 dB azimuth beamwidth of the array is relatively constant with respect to the azimuth bearing but varies with the elevation bearing according to

$$\phi_{BW} = 7.8°/|\cos\psi| \quad \text{for } |\psi| < 90° \tag{174}$$

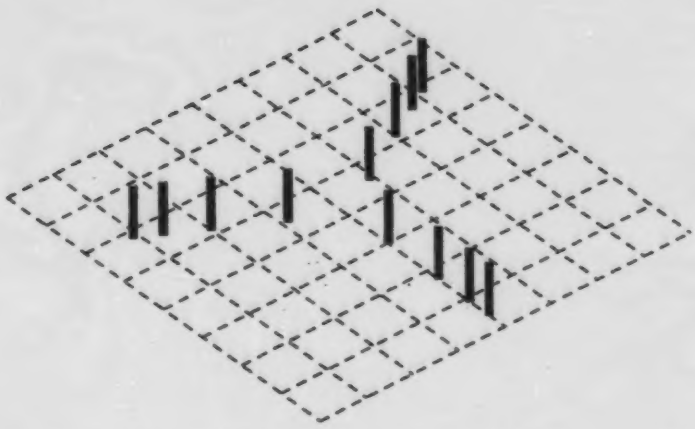


Figure 6: Three dimensional view of the antenna array. Each grid square has a dimension of $1\lambda \times 1\lambda$.

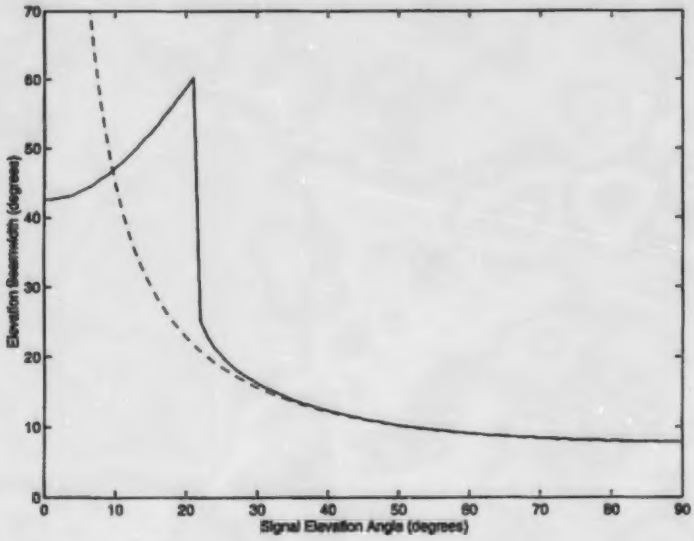


Figure 7: Antenna array elevation beamwidth as a function of elevation comparing the simulated response (solid line) with the $\sin^{-1}\psi$ predicted response (dashed line).

for the array size shown. Note that measuring the azimuth beamwidth at or near $\psi = 90°$ is effectively meaningless. The 3 dB elevation beamwidth is given by

$$\psi_{BW} = 7.8°/|\sin\psi| \quad \text{for } |\psi| > 30°. \qquad (175)$$

For elevation bearings below 30° the 3dB beamwidth is somewhat more complicated as shown in Figure 7. The failure of the above expression at the lower elevation angles is due to the distortion of the main lobe in the antenna pattern at the elevation angle ψ by the reflection of this lobe at the elevation angle $-\psi$. For example a 2-dimensional x-y array, using the free space assumptions, has a symmetrical gain pattern for elevation angles above and below the horizon, i.e., a main lobe and a reflection lobe. At low signal elevations, the main and reflection lobes begin to merge. Using the 3 dB criteria, the lobes are considered merged when the minimum gain between the two lobes is greater than -3 dB the maximum gain. For the array configuration shown this occurs at 21.2°. At this point, the beamwidth effectively doubles. For even lower signal elevations the merged lobes move closer together so the beamwidth actually decreases.

When processing the simulated data using the SML algorithm, the default values for the parameters under investigation (unless otherwise specified) were: a grid spacing of 0.3 times the 3 dB beamwidth of the antenna array ($d_\phi = 0.3\phi_{BW}$ and $d_\psi = 0.3\psi_{BW}$); a data block size of $K = 100$ samples; the raised-cosine elliptical taper function; and an assumed number of signals equal to the actual number of signals.

The processed results were quantified in two ways. The first was the measurement of the failure rate of signal bearing estimates, and the second was the measurement of the accuracy of the successful estimates. In this report, a bearing estimate was considered to be a failure if it deviated from the true signal bearing by more than half the array beamwidth (taking into account both the azimuth and elevation beamwidths). Accuracy was determined by calculating the root-mean-squared (RMS) error of the estimates according to

$$RMS\ Error = \sqrt{\frac{1}{H}\sum(\hat{\phi}_m - \phi_m)^2 + (\hat{\psi}_m - \psi_m)^2} \qquad (176)$$

where the summation was performed for all H successful estimates of signal m. Other specifics of the simulations and processing were dependent on the aspect of the SML algorithm being investigated, and are discussed in the following three sections.

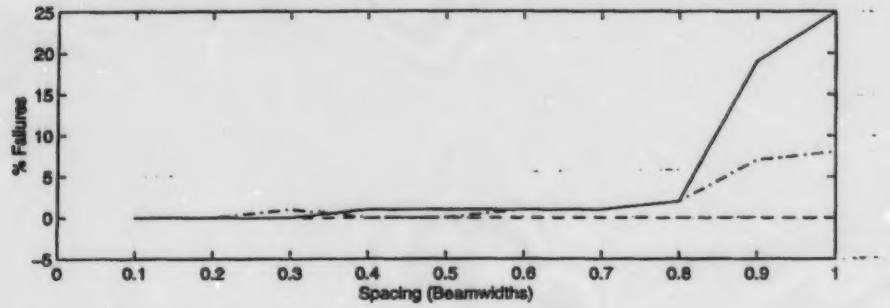
6.1 Signal Model Grid Spacing

The angular spacing between point-source signals used in the signal model grid must be sufficiently narrow for the model to adequately represent a real spread signal. However, making this spacing too narrow can unnecessarily increase the number of computations involved in generating these models (the increase in computations is inversely proportional to the square of the spacing). Hence it is useful to determine the widest spacing that can be used before introducing too much error due to the poorer modeling that results.

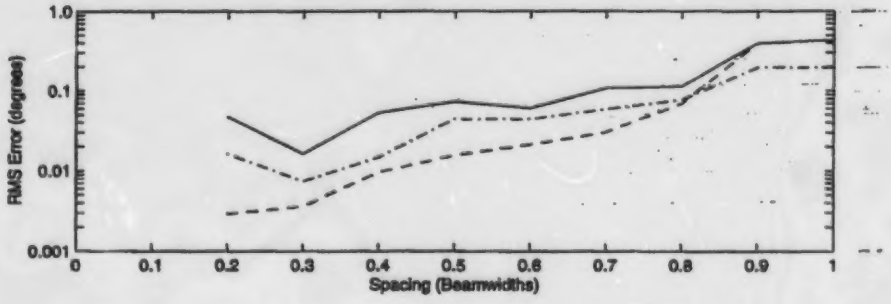
Since the resolving power of any antenna array is a function of the beamwidth of the array, a natural choice would be to use a spacing value which is some fraction of the beamwidth. To evaluate the effect of changing the azimuth and elevation grid spacing values, a series of 100 simulations were carried out generating 100 sets of data. The data set was then processed using grid spacings which were varied from 0.1 beamwidths up to 1.0 beamwidths in 0.1 beamwidth increments in both azimuth and elevation relative to the appropriate azimuth and elevation beamwidth. The results are shown in Figure 8.

In generating Figure 8 the true bearings were taken to be those estimated using the 0.1 beamwidth spacing instead of the actual bearings given in Table 1. This is the only occasion that it was done this way. The intent was to find the widest spacing that yielded the same results (or reasonably close to the same results) as if the signal model had been generated using infinitesimally narrow spacings. A spacing of 0.1 beamwidths was considered sufficiently narrow for these purposes. Note also that perturbations observed in the plotted results (relative to the general trends) are almost certainly statistical effects introduced by using finite (versus infinite) data sets.

Examining the relative failure rate, it remains low up until 0.8 beamwidths. (Using the actual signal bearings from Table 1, instead of the values estimated for a spacing of 0.1 beamwidths, the failure rate varies from 18% to 20% for the same beamwidths). Above 0.8 beamwidths the increase is more dramatic. From these results it would appear that spacings of up to 0.8 beamwidths could be used with reasonable performance still achieved. For the rest of the results provided in this report a more conservative spacing of 0.3 beamwidths was used.



(a)



(b)

Figure 8: Effect of grid spacing on SML estimation performance relative to the performance for a grid spacing of 0.1 beamwidths. The plots show (a) the relative failure rate for each of the three bearing estimates, and (b) the corresponding relative accuracy of the successful estimates. The solid lines, dashed lines, and dash-dot lines correspond to the results for the point-source signal at $\phi = 50°$, the spread-source signal at $\phi = 180°$, and the spread-source signal at $\phi = 210°$, respectively.

6.2 Data Block Size

In the assessment of the effects of ionospheric scattering patches, it has been found that signals arriving from one part of a scattering region appear to be uncorrelated with signals arriving from any other part of the scattering region. To take advantage of this requires collecting enough samples of sensor data over a long enough time period that the signals become reasonably decorrelated. Assuming that the samples are taken at intervals which exceed the correlation time constant of the data, then the question becomes: how many samples are enough?

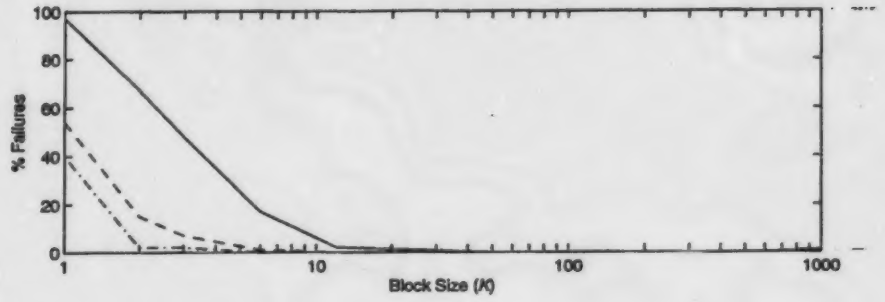
To answer this question, a series of simulation data sets were produced. Each set consisted of 100 trials with a given number of sensor samples (or block size) K collected for each trial, which was varied from set to set. To account for the most obvious effects of averaging inherent in processing larger block sizes (which leads to an equivalent increase or processing gain of $10 \log K$ dB in the SNR), the noise level was also adjusted according to

$$\textit{Noise Level} \;=\; 10 \log K \;-\; 35 \text{ dB}. \tag{177}$$

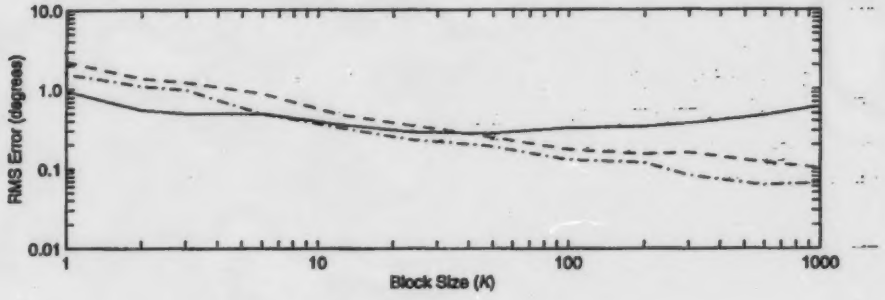
This noise level compensation was done to avoid masking other effects which might not otherwise be apparent. The results are shown in Figure 9.

Inspection of the failure rate reveals that it is high for very small block sizes (e.g., $K = 1, 2, 3$) but decreases rapidly to zero as the block size increases. The accuracy results are interesting. Ignoring the effects of simple averaging, then for the point-source signal, there is no significant change in accuracy as the block size increases since it can be well modeled, even for a single sample. Additionally it is already spatially decorrelated from the other two signals given the angular separation of several beamwidths. The bearing estimation accuracy for the spread signals decreases at a relatively constant rate with a reduction in the error by a factor of 10 as the block size is increased from 1 to 100. The improvement is a reflection of the fact that the larger block sizes enhance the effects of decorrelation on which the signal modeling relies. Given that the spread signals are ten times greater in power than the point-source signal, this improvement would only be expected to continue until the RMS errors were on the order of ten times less than that of the point-source.

This upper limit on accuracy as a function of block size is confirmed in the simulation results shown in Figure 10. In this case only the two spread signals were simulated and the noise level was 15 dB higher than specified in (177). Comparing the RMS error results

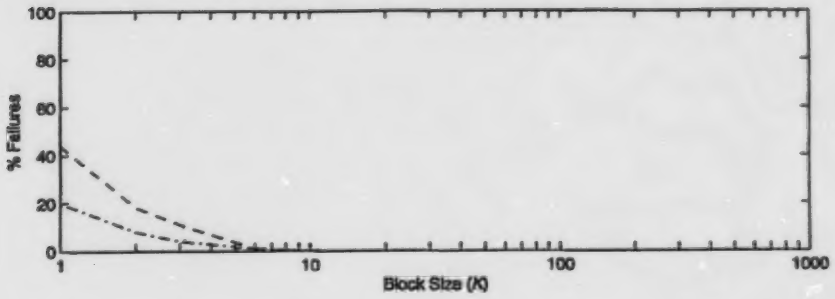


(a)

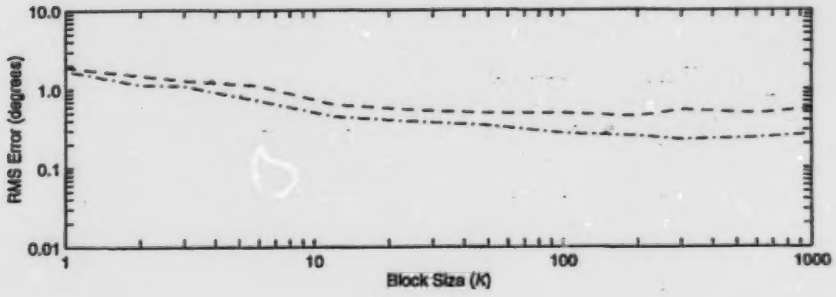


(b)

Figure 9: Effect of the number of data samples (block size) on SML estimation performance showing (a) the failure rate for each of the three bearing estimates, and (b) the relative accuracy of the successful estimates. The solid lines, dashed lines, and dash-dot lines represent the point-source signal at $\phi = 50°$, the spread-source signal at $\phi = 180°$, and the spread-source signal at $\phi = 210°$, respectively. Note that the effect of simple averaging as a function of block size has been removed.



(a)



(b)

Figure 10: Effect of the number of data samples (block size) on the estimation of the spread signals for a lower SNR (15 dB lower than in the previous example) showing (a) the failure rate for the two spread signal bearing estimates, and (b) the relative accuracy of the successful estimates. The dashed and dash-dot lines represent the spread-source signal at $\phi = 180°$ and the spread-source signal at $\phi = 210°$, respectively. Note that the effect of simple averaging as a function of block size has been removed.

to those in Figure 9(a), below $K = 12$ the results are very similar despite the difference in SNR. Above this, the RMS error plots in Figure 10 flattens out as suggested already, indicating that sufficient decorrelation has occurred for the spread signals to be accurately modeled. For these higher values of K, the dominant error mechanism is noise.

The total effect of increasing the block size on RMS bearing error can be determined by including the effects of simple averaging. The effect of averaging on accuracy is found by considering the equivalent increase in SNR and then utilizing the results from Section 7.1 to relate this to the RMS bearing error. Combining this with the results observed here, then the block size can be related to the RMS error ϵ according to

$$\epsilon \propto \begin{cases} K & \text{for small K} \\ \\ \sqrt{K} & \text{for large K} \end{cases} \tag{178}$$

where a constant SNR is assumed. In the above expression, small K means that not enough samples are available for sufficient signal decorrelation, while large K means there are enough samples and noise effects dominate. For a point-source signal the delineation between small and large is $K = 1$.

Based on the failure rate results, $K = N$ (number of samples equal to the number of sensors) is a reasonable lower limit for the number of samples. A more suitable choice would be to select a value of K large enough that sufficient signal decorrelation has occurred, although this would involve a careful assessment of the expected signal environment in terms of the expected signal SNR's and spreading values. For the rest of the simulations that follow in this report $K = 100$ was deemed suitable.

One final observation is that wider spatial spreading of the received signal results in poorer RMS accuracy of the bearing estimate (everything else being equal). This effect is discussed in more detail in Section 7.2 .

6.3 Taper Function

The choice of taper function is ideally based on the shape of the ionospheric scattering region and the density of signal power across this region. However, since the ionosphere is highly dynamic, it is likely that there will be differences in these characteristics between different scattering regions, or even differences for the same region observed at different

periods of time.

To investigate the effect of mismatches between the assumed signal model and the actual signal spreading distribution, three simulation data sets were produced, each consisting of 100 trials. The first data set was produced using the signal parameters listed in Table 1 and using the raised-cosine elliptical power distribution to generate the simulated signals. The second data set was produced using the signal parameters listed below in Table 2 and used the uniform rectangular power distribution to generate the simulated signals. The third data set was produced also using the signal parameters listed below in Table 2, but using the asymmetric power distribution shown in Figure 11 to generate the simulated signals. For all simulations the SNR was set to -12 dB (lowered from the default value of -15 dB to increase the probability of failures).

Table 2: Simulation Signal Parameters for Uniform Rectangular and Asymmetric Spread Distributions

Signal	ϕ	ψ	Δ_ϕ	Δ_ψ	Power
1	50°	10°	0°	0°	-20 dB
2	180°	30°	18.9°	6.3°	0 dB
3	210°	25°	9.5°	3.2°	0 dB
noise	–	–	–	–	-12 dB

To simplify the comparison of results, the spread values in Table 2 were chosen so that processing any of the data sets using the same SML algorithm variant produces nearly the same mean spread values. For example, using the SML algorithm with the raised-cosine elliptical taper function (denoted SML(e)) produced mean spread values that were the same as those listed in Table 1 for all three data sets. This means that the spread results were correct for the first data set but were overestimated for the other two data sets the results due to the modeling mismatch. Similarly, using the SML algorithm with the uniform rectangular taper function (denoted SML(u)) produced mean spread values that were the same as those listed in Table 2 for all three data sets. In this case the spread results for the first data set were underestimated while the results for the other two data sets the results were correct.

To process the three data sets, both the SML(e) and SML(u) algorithm variant were employed. For both variants grid spacings of $d_{\phi_m} = 0.1\phi_{BW}$ and $d_{\psi_m} = 0.1\psi_{BW}$ beamwidths were used because it was found that for spacings of 0.3 beamwidths the failure rate significantly increased when using SML(u) (but not SML(e)). Table 3 shows the overall RMS bearing error and failure results for each signal. The results are shown in order,

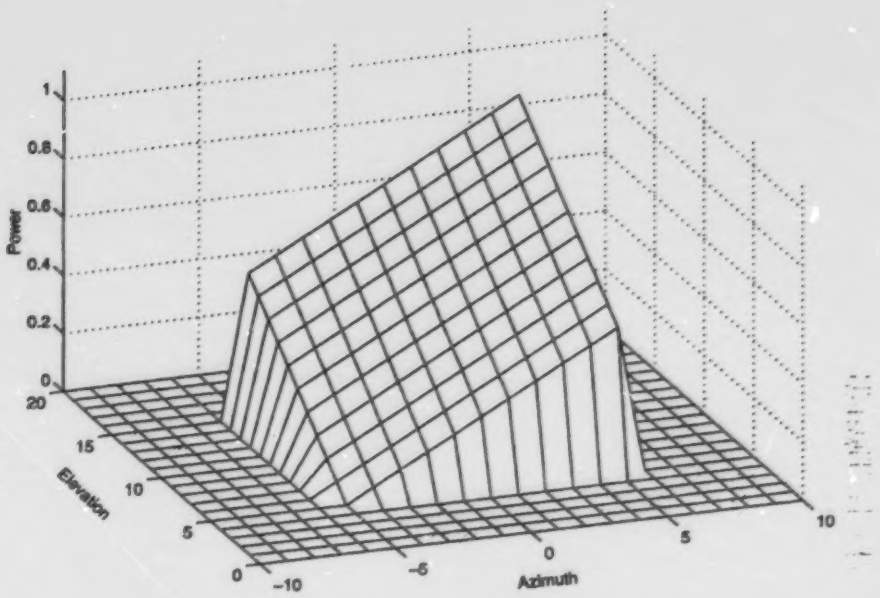


Figure 11: Asymmetric power distribution used for spread signal simulations.

that is, the corresponding bearing angles are $\phi = (50°, 180°, 210°)$ and $\psi = (10°, 30°, 25°)$.

Table 3: Effect of Taper Function on SML Performance

	SML(e) Results		SML(u) Results	
Signal Type	RMS Errors	Failure Rate	RMS Errors	Failure Rate
Elliptical	(0.50°, 0.20°, 0.14°)	(1%, 0%, 0%)	(0.54°, 0.20°, 0.15°)	(5%, 0%, 0%)
Uniform	(0.41°, 0.17°, 0.13°)	(2%, 0%, 0%)	(0.42°, 0.16°, 0.12°)	(2%, 0%, 0%)
Asymmetric	(0.42°, 0.95°, 0.53°)	(1%, 0%, 0%)	(0.44°, 0.77°, 0.47°)	(2%, 0%, 0%)

Examining the results for the point-source (at $\phi = 50°$), SML(e) performed better than SML(u) regardless of the spread distribution types of the other two signals. For the spread signals (at $\phi = 180°$ and 210°), the SML algorithm employing the matched taper function performed better for both spread signal types (i.e., SML(e) performed best for signals using the raised-cosine elliptical spread distribution, and SML(u) performed best for signals using the uniform rectangular spread distribution). The degradation that resulted from using an improperly matched taper function was almost insignificant in terms of accuracy for the uniform and elliptical signal types although there was an increase in the failure rate for SML(u). For the asymmetric signal type, the accuracy in estimating the spread signals was significantly degraded although the errors were still less

than 1° . The poorer accuracy was due to the signal models for both SML(e) and SML(u) being biased towards the peak in the asymmetric spatial power distribution of the spread signals. This biasing was greater for SML(e) than SML(u). Removing the biasing yields error results very similar to the results for the elliptical and uniform signal types.

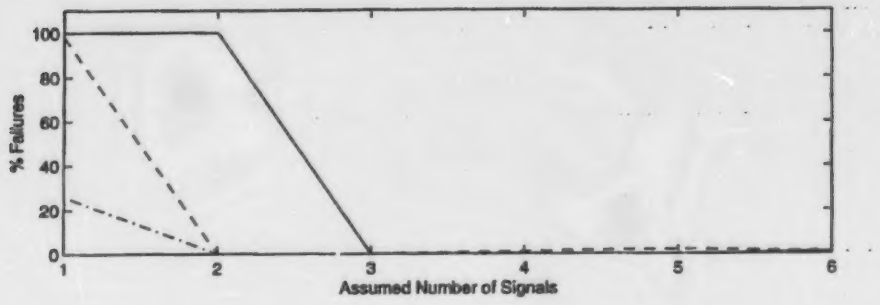
Based on preceding results, it seems clear that the best performance, not surprisingly, is obtained when the taper function is matched to the actual signal spread distributions, although the SML approach seems to be reasonably robust when mismatches occur (at least for the type of array tested in this report – a larger array with a narrower beamwidth could lead to a greater sensitivity to modeling errors). Given the superior performance of the raised-cosine elliptical taper function for point-source signals, and the greater likelihood that this taper function would be better matched to real world signals, it was used exclusively throughout the rest of this report.

6.4 Assumed Number of Signals

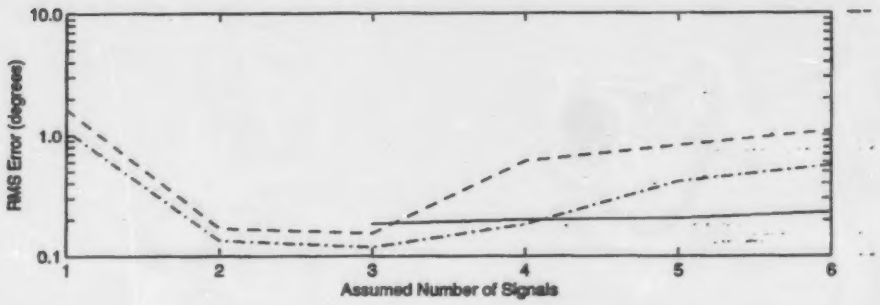
A common difficulty with many advanced direction finding algorithms is determining the correct number of signals intercepted in the receiver pass band. In this report, unless otherwise stated, it is assumed that the number of signals is known. Given that the number of signals may sometimes be incorrectly chosen, it is useful to know what effect this has on performance of the SML algorithm. Using the default settings for the simulations (as discussed in Section 6.0) to produce 100 trials, the data was then processed assuming from one signal up to six signals. The overall results are shown in Figure 12.

Inspecting these results reveals several things. When the number of signals is underestimated, then either the strongest signals are detected only, or two or more signals are wrongly identified as being a single signal. For example, in the case where only one signal is assumed, the weakest signal (the point-source signal at $\phi = 50°$) is never seen. The spread signal at $\phi = 210°$ is sometimes correctly identified while the rest of the time the two spread signals are mistakenly identified as a single signal with an intermediate bearing. In terms of initial detection, the spread signal $\phi = 210°$ is favoured over the spread signal at $\phi = 180°$ since, even though the two signals have equal power, the signal at $\phi = 210°$ has less spreading giving it a higher power density (more power per solid angle). When two signals are assumed, the two stronger spread signals are properly estimated, but the weaker point-source signal still remains undetected.

When the number of signals is overestimated, either spurious results are generated, or



(a)



(b)

Figure 12: Effect of the assumed number of signals on SML estimation performance showing (a) the failure rate for each of the three bearing estimates, and (b) the relative accuracy of the successful estimates. The solid lines, dashed lines, and dash-dot lines represent the point-source signal at $\phi = 50°$, the spread-source signal at $\phi = 180°$, and the spread-source signal at $\phi = 210°$, respectively. The results for spurious signals are not shown.

a single spread-source signal is identified as two spread-source signals. The splitting of the spread-source signals explains the worsening error as the number of assumed signals is increased past three (the bearings of the split signals tend to be offset on either side of the true signal).

In general, from the point of view of estimator performance, it is better to overestimate the number of signals than underestimate the number of signals, since overestimating yields all the signal directions. The main problem with overestimating is the generation of spurious signal directions and errors due to splitting of the spread region.

7.0 SML PERFORMANCE

The previous section dealt with modifying various control parameters and the effect these modifications had on the performance of the SML algorithm. In this section, the effect of various uncontrolled parameters and their effect on the performance of the SML algorithm is studied. These parameters are discussed in the next few sections and include SNR, signal spreading, and angular spacing between signals. Where appropriate, the results using the MUSIC DF algorithm are also shown for comparison purposes.

7.1 Effect of Noise

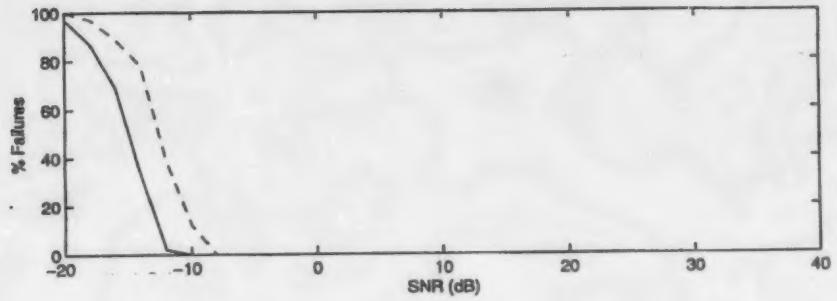
As was done when investigating the effects of various control parameters, the effect of noise was investigated through simulation. In the first series of simulations, a single point source signal at $(\phi, \psi) = (180°, 30°)$ was generated and the SNR was varied from -20 dB to +40 dB in 2 dB increments. One hundred trials were generated for each SNR setting. In the second series of simulations, the point-source signal was changed to a spread-source signal with spread parameters $(\Delta_\phi, \Delta_\psi) = (30°, 15°)$, but all other parameters remained the same. The results are shown in Figure 13.

Several features of the results are worth pointing out. The SNR at which the failure rate dramatically departs from 0% is called the threshold SNR. For the point-source this occurs between -12 and -10 dB and for the spread-source, it is slightly worse, occurring between -8 and -10 dB. The poorer performance for the spread signal (including both threshold and accuracy) is a function of the amount of spreading and is discussed in more detail in the next section.

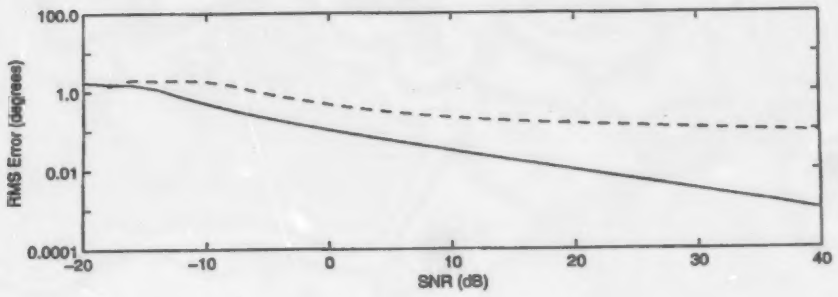
The scales used to display the RMS bearing errors in Figure 13(b) were chosen because they linearize the accuracy results for the point-source signal above the threshold SNR. In this region, for every 20 dB increase in SNR the RMS bearing error is reduced by a factor of 10. Written mathematically, the relationship is expressed as

$$\epsilon \propto \sqrt{SNR} \tag{179}$$

where ϵ is the RMS bearing error. The spread-source also exhibits the same behaviour between -8 and 0 dB, but above this SNR the error begins to level out as the uncertainty in the signal model begins to dominate the error (as discussed previously in Section 6.2).



(a)



(b)

Figure 13: Effect of the signal-to-noise level on SML estimation performance showing (a) the failure rate, and (b) the accuracy of the successful estimates. The solid lines represent the point-source signal and the dashed lines represent the spread-source signal.

At higher SNR's then, the only way to improve accuracy would be to use a larger block size assuming the signal was well modeled.

To determine whether including the azimuth and elevation spread parameters in the SML signal model degrades performance when estimating point-source signals, the point-source simulation results were reprocessed using the MUSIC algorithm. The results were virtually identical to the SML results indicating that any degradation that does occur is insignificant.

7.2 Effect of Signal Spreading

The effect of signal spreading was investigated through simulation by measuring accuracy as the azimuth and elevation spreads of the generated signal was varied. In the simulation a single signal at $(\phi, \psi) = (180°, 30°)$ was used with three different SNR's, namely, 0, 15, and 30 dB. The azimuth spread was varied from 1° to 30° according to the sequence $\Delta_\phi = 1°, 2°, 4°, 6°, 8°, 10°, 15°, 20°, 25°, 30°$ for each value of the SNR. The elevation spread was set to one half the azimuth spread ($\Delta_\psi = \Delta_\phi/2$). One hundred trials were generated for each spread and SNR setting. The results are shown in Figure 14.

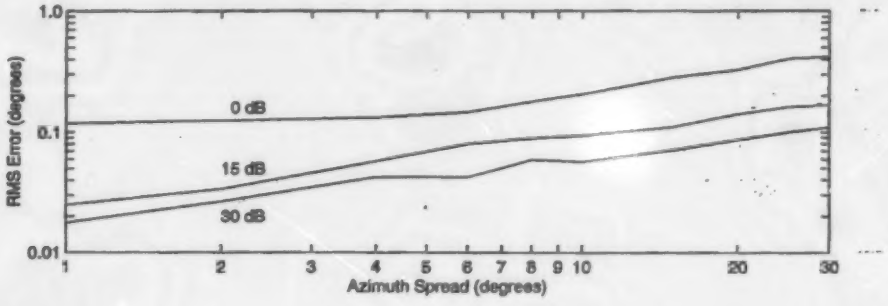


Figure 14: Effect of signal spread on SML accuracy for SNR's of 0, 15, and 30 dB. Elevation spread was varied according to one half the azimuth spread.

The effect of signal spreading on accuracy can be attributed to two factors: signal model uncertainty and the filter effect. The first factor, signal model uncertainty arises from the fact that the spatial model for the spread signals is stochastic and requires a sufficient number of snapshots to build up the appropriate statistics in the data as was discussed in Section 6.2. For a single point-source signal, a single snapshot is sufficient, but for a spread-source, the number of snapshots required to achieve a given accuracy rises as the spread region increases in size (e.g., compare the accuracies of the two spread

signals in Figure 9(b)). Conversely, for a given number of snapshots, accuracy degrades as the spread region increases.

The effect of the model uncertainty can be seen in the accuracy results for the SNR = 30 dB setting. Compared to the accuracy results for the SNR = 15 dB setting, the improvement at the higher SNR is only marginal (e.g., compare this to the improvement that occurs when the SNR is increased from 0 to 15 dB). Hence the main source of error at the higher SNR is model uncertainty, even for spreading as narrow as 1°. For example, from Figure 14 the accuracy for $\Delta_\phi = 1°$ is approximately 0.02° RMS while the accuracy for a point-source signal under the same conditions is approximately 0.003° RMS (see Figure 13).

The second factor, the filter effect, arises from the fact that many advanced DF estimators, such as MUSIC and SML, can be interpreted as techniques which use spatial filters to suppress the signal content of the input data in the estimation procedure (consider the discussion of the whitening filter in Section 3.2). For point-source signals, this requires creating a filter with M notches corresponding to the M signals. These notches are then adjusted so they are located in the same direction as the signals themselves.

The adjustment of the notch can be made by measuring the filtered power and then adjusting each notch to minimize this power. Ideally the beamwidth of each notch would be infinitely narrow so that the filtered power would be reduced only when the notch was placed exactly in the direction of a signal. In reality, the width of the notch is finite so that some noise also passes through the filter which, in turn, affects the accuracy of the adjustment procedure. The width of the notches is related to the inverse of the array aperture, hence an array with a smaller aperture provides better accuracy.

For spread signals, the beamwidths of the generated filters become much wider because they are designed to match the spread widths of the signals. The wider beamwidth naturally results in a greater amount of noise passing through the filter with a correspondingly greater degradation in accuracy. Given that the results for SNR = 30 dB are dominated by the model uncertainty, the results for SNR = 0 dB are certainly dominated by the filter effect. In this case, there is little degradation in accuracy for spreading angles less than $4 - 5°$ (one half the beamwidth of the array) compared to the accuracy of a point-source signal (0.1° RMS). This is due to the limitations on the narrowness of the filter beamwidths discussed above. For greater spreading values, the change in accuracy due to noise as a function of spreading angle follows approximately the same trend as for the change in accuracy due to model uncertainty.

7.3 Detection of a Weaker Signal

As mentioned in the introduction, it is important that a DF algorithm be able to detect a weaker point-source signal in the presence of stronger spread-source signals. To evaluate the performance of the SML algorithm, simulations were run involving a single spread-source with a fixed bearing and a single point-source whose bearing was adjusted incrementally, beginning with a large initial angular difference, until the two bearings coincided. After each increment, the signal power of the point-source was increased from a low value in 0.5 dB intervals until the failure rate dropped below 5% (5 out of 100 trials). The corresponding SNR of the point-source signal is defined here as the threshold SNR and provides a good indication of the limits of detectability of the point-source signal for the given signal environment. The relevant signal and noise parameters are shown in Table 4.

Table 4: Signal Parameters for Signal Detectability Simulation

Signal	ϕ	ψ	Δ_ϕ	Δ_ψ	Power
1	180°	30°	30°	15°	0 dB
2	adjusted	30°	0°	0°	adjusted
noise	–	–	–	–	-20 dB

The results from the simulations are shown in Figure 15. In this case the failure rate is not shown since it was fixed to approximately 5%. From the results it is clear that when the point-source and spread-source signals are well separated, 25° or more ($> 1.2\phi_{BW}$), the presence of the spread source signal has a small but significant effect. For example, the threshold SNR for a point-source without any other signals present occurs at -12 dB (see Figure 13). In the presence of the spread-source, the threshold SNR ranges from -10 to -7 dB which is a degradation of 2 to 5 dB. As the separation between signals is reduced from 25° to 0°, the threshold SNR increases dramatically until it reaches a maximum at 16 dB. However, even with both signals arriving from coincident directions, detection of both signals is not only possible, but the minimum point-source power is still 4 dB less than the spread-source power.

The differences between the wide and narrow separation cases can be understood in terms using a spatial filter to suppress the effects of the spread-source signal. In the wide separation case, the filtering can be accomplished easily, leaving only the noise as the main source of error. In the narrow separation case, it is more and more difficult to filter out the spread signal independently of the point-source signal as the separation decreases.

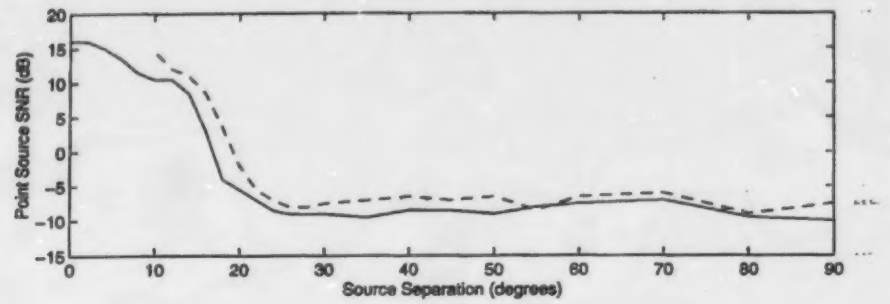
Consequently, the spread-source signal begins to act as a strong noise background and the threshold increases accordingly.

The accuracy results for the two signals are relatively constant for wide spacing and relatively independent of the presence of each other. For example, in the single signal case, the accuracies were measured to be 0.15° for the spread-source and no point source, and 0.7° for the point-source and no spread source – worse for the point-source due to the much lower SNR. For narrower spacings, the threshold signal power of the point-source signal increases, degrading the accuracy of the spread-source signal until the separation becomes as small as 8°. Under 8° separation, the accuracy actually improves even though the signal power of the point-source signal is still increasing, perhaps because the two signal directions are too similar to cause an appreciable error. The accuracy of the point-source remains in the range of $0.5° - 1.5°$ for most separation angles with the lowest errors occurring for the smallest separations.

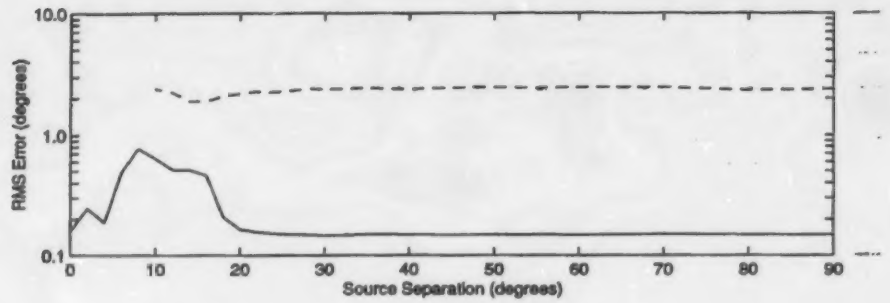
The simulations were also repeated using the MUSIC algorithm. The assumed number of signal directions (a parameter also required by this algorithm) used was six, since occasionally up to five directions were required to describe the spread-source leaving at least one direction for the point-source. In most cases, however, *only three or four signal* directions were required for the spread signal, resulting in one or more false direction estimates. This is illustrated in Figure 16 which shows the results of processing 100 trials using MUSIC.

Generally, the necessity of using several signal directions to describe a spread-source and the problem of extraneous signals, makes the interpretation of the MUSIC results somewhat problematic. For the sake of this report, only the two (out of six) estimated signal bearings closest to the true signal bearings were used when generating the statistical results. Additionally, since for separation angles less than 10° there was no obvious way to determine whether a peak was associated with the spread signal or the point-source signal (compare the examples shown in Figure 17(a) and (b)), no statistical results for MUSIC were calculated for these angular separations. The failure and accuracy results are shown as the dashed lines in Figure 15.

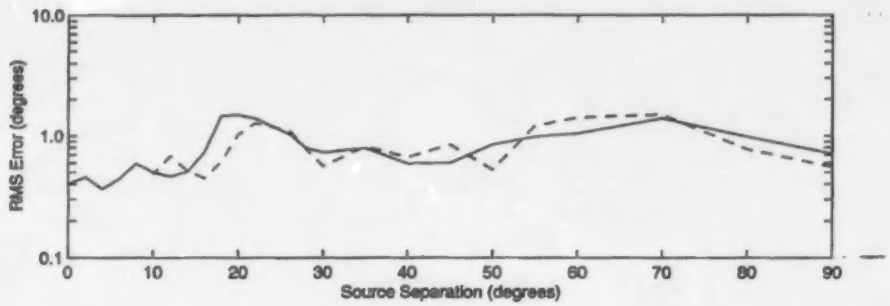
Comparing the results for the SML and MUSIC algorithms, for all signal separations measured, SML performed better. In terms of threshold performance, the differences were no more than 2.5 dB for the wider signal separations, but are as large as 7.5 dB for the narrower separations. Additionally, for angular separations less than 10°, the SML algorithm was still able to produce results which could be unambiguously associated with



(a)



(b)



(c)

Figure 15: Effect of angular spacing on the ability to detect a weaker point-source signal in the presence of a stronger spread-source signal showing (a) the detection threshold SNR for the weaker point-source signal, (b) accuracy of the spread-source estimates at threshold, and (c) accuracy of the point-source estimates at threshold. The solid lines represent the SML results, and the dashed lines represent the MUSIC results.

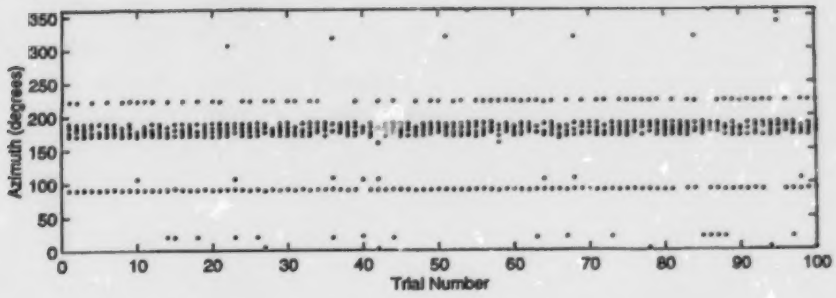


Figure 16: MUSIC azimuth bearing estimates for 100 trials assuming 6 signal directions. Bearing estimates of the spread-source signal at 180° and the point source signal at 90° are clearly evident along with various false signal bearing estimates (particularly at 222°).

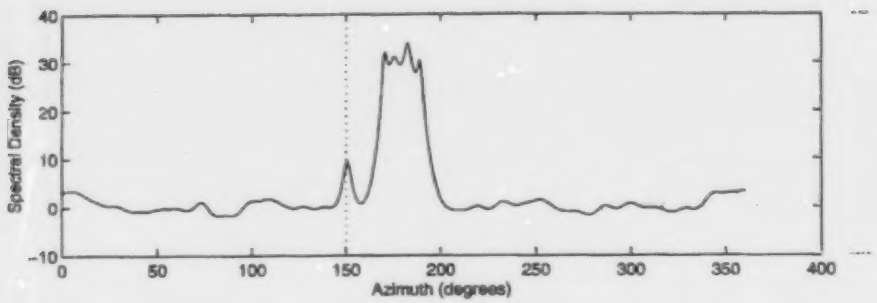
the two signals using either signal power or spread estimates, whereas with the MUSIC algorithm this was not possible as previously discussed.

The accuracy in estimating the point-source signal direction was the same for both algorithms. However, since the SML algorithm has a lower threshold, this implies that tested at the same SNR (e.g., the MUSIC threshold) the SML algorithm is more accurate than MUSIC. The accuracy of MUSIC for estimating the direction of the spread-source signal was very poor, highlighting the difficulty of estimating spread-source signals using a point-source model.

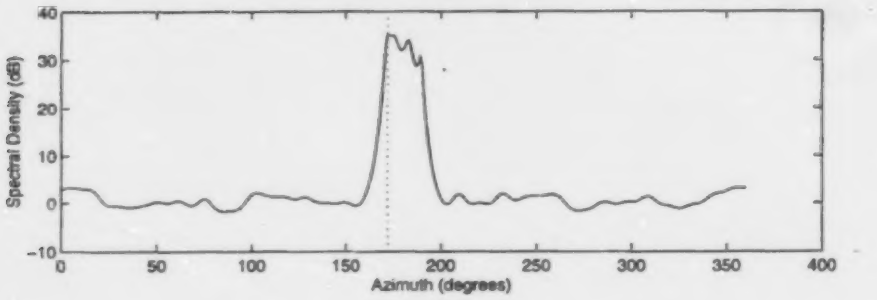
Generally the results show that for signal environments with spread-source signals, improved modeling leads to significantly better performance.

7.4 Performance using High Latitude Off-Air Data

The ultimate test for the SML algorithm is against the data for which it was originally developed, namely actual high latitude HF measurement data. To this end, an off-air measurement data set was processed which had been collected on January 23, 1996, using the Vortex DF system at CFB Alert which is located on the northern tip of Ellesmere Island in Northern Canada (82.50° N, 62.35° W). The received signal was from the CZD transmitter located further south at Iqaluit (63.45° N, 68.30° W) and transmitting at a frequency of 9.292 Mhz. The great circle signal bearing (azimuth) was 188.5° measured clockwise from north. The signal itself was composed of a 15 second tone followed by a 15 second Morse code call sign repeated continuously for a total period of 25 minutes. This was, in turn, followed by a 5 minute sounder slot during which time there was no



(a)



(b)

Figure 17: Detection of a weaker point-source signal in the presence of a stronger spread-source signal using MUSIC. The azimuth spectrum is shown for an elevation angle of $\psi = 30°$. The simulation parameters for the noise and spread-source are listed in Table 4, and the point-source parameters were (a) $\phi = 150°$ with a signal power of -7.5 dB and (b) $\phi = 172°$ with a signal power of 14.5 dB. When the separation between the spread-source and point-source is too narrow, as in (b), there is confusion as to which spectral peak belongs to which signal.

transmission. This whole sequence was then repeated over and over again.

The Vortex system consisted of an array of 12 antennas connected to a 12 channel receiver. The antenna array arrangement, which is as shown in Figure 18, utilized 8 elevated feed monopole antennas from the inner ring of a Pusher array (a circular array with 24 antennas) and 4 outlying antennas (also elevated feed monopole antennas) used to increase the array aperture.

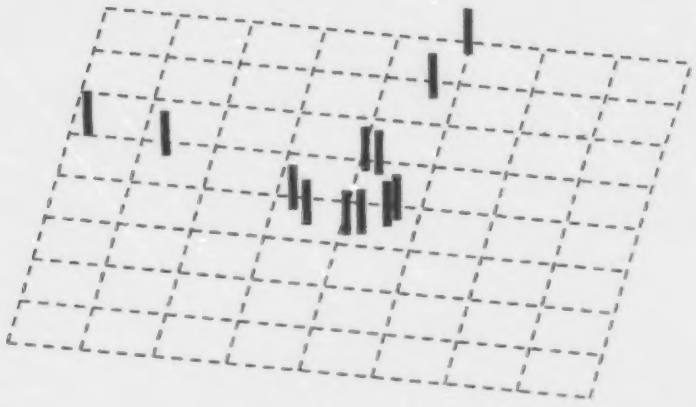


Figure 18: Three dimensional view of the Vortex antenna array. Each grid square has a dimension of $1\lambda \times 1\lambda$.

The Vortex receivers were used to downconvert the input signals from HF to 2.5 kHz with a filtered bandwidth of 3.5 kHz. The downconverted signals were then digitized at a rate of 10 kHz.

To generate the covariance estimates, an FFT was first performed on each block of *12 × 16000 data points* and only the positive frequency data corresponding to 3.8 to 4.2 kHz were retained. This served the dual purpose of converting the data to IQ format and suppressing interference due to noise and other unintended HF signals. The data covariance matrix was then formed directly from this frequency domain data.

The particular time of the data collection was a very active period in terms of ionospheric disturbances. Previous analysis using the MUSIC algorithm [9] found that there were very large bearing swings over time. A high degree of scatter in the measured bearings was also observed, which is indicative of spread signals. The results of reprocessing

the data as described here and using the MUSIC algorithm are shown in Figure 19. For this processing, the number of signals parameter was set to 7 which was very likely an overestimate of the number of signal directions required, but which was far less detrimental to the overall accuracy than underestimating this number would have been. For display purposes, only the bearings corresponding to the four largest peaks in the MUSIC spectrum are plotted in the figure. Additionally, thresholding was applied to reject peaks in the MUSIC spectrum which were clearly due to noise.

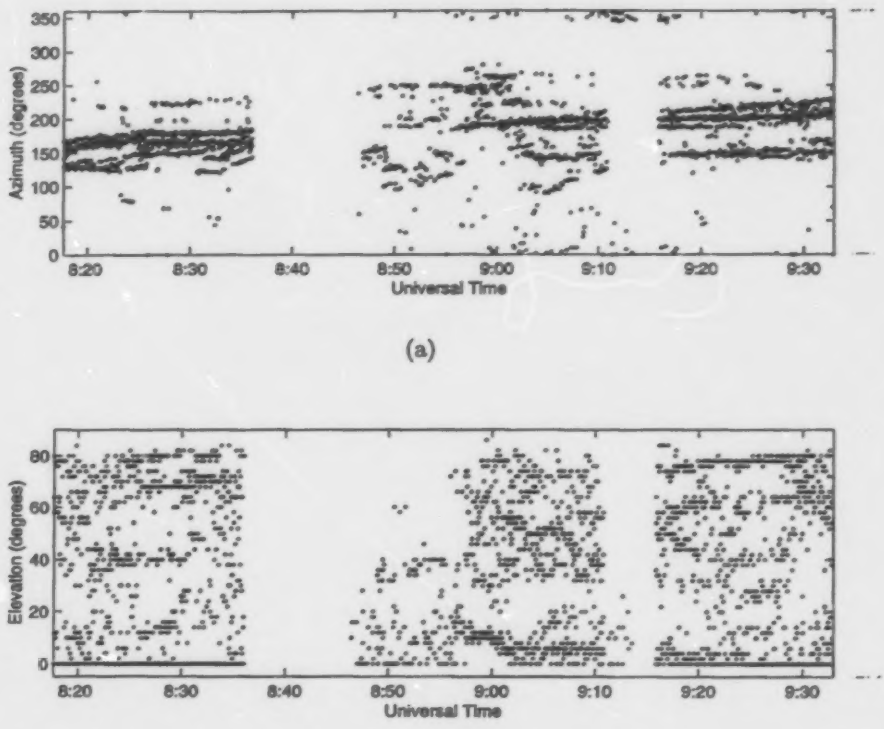
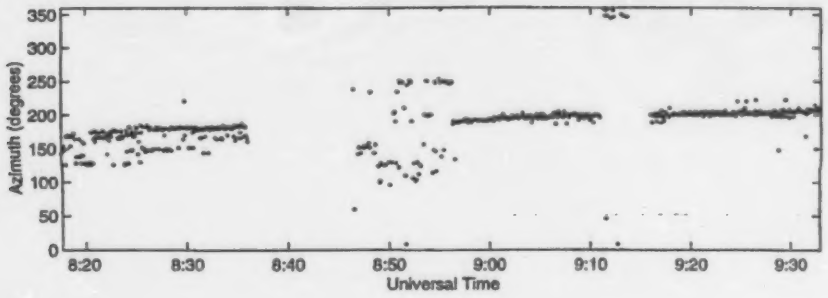


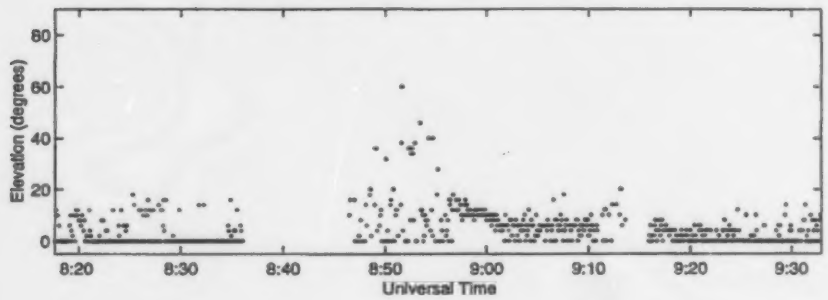
Figure 19: MUSIC results for high latitude off-air data collected on January 23rd 1996. Only the directions corresponding to the four highest peaks in the MUSIC spectrum are shown for (a) azimuth, and (b) elevation.

Even limiting the number of signal directions displayed leads to some confusion in the interpretation of the results, particularly in elevation. To provide some additional insight, the results have been replotted in Figure 20, except this time showing only the direction of the signal with the largest peak. It is quite evident from these results that the azimuth

bearing was increasing over time and that there was considerable scatter in the results.



(a)

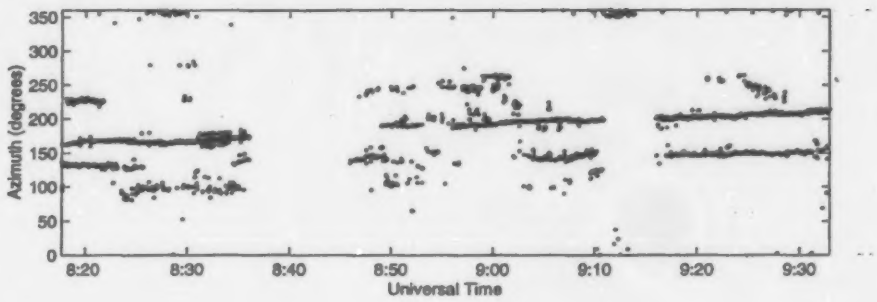


(b)

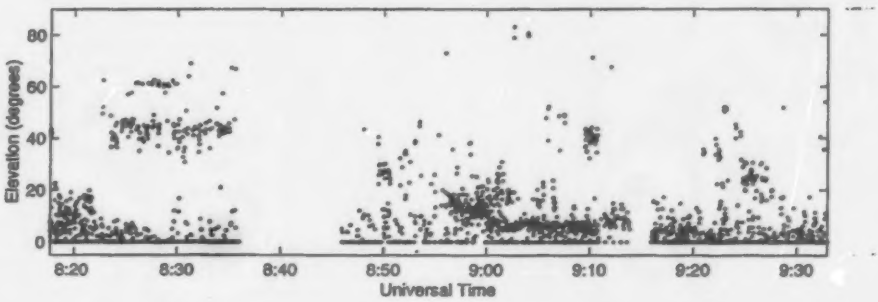
Figure 20: MUSIC results for high latitude off-air data collected on January 23rd 1996. Only the direction corresponding to the largest peak in the MUSIC spectrum is shown for (a) azimuth, and (b) elevation.

Before discussing the SML results, a comment is in order about the block size. Although a block size of 16000 (representing 1.6 seconds) would appear to be sufficient for SML processing purposes, this does not take into consideration the requirement for sample to sample decorrelation. The time required to achieve this decorrelation is related to the Doppler spreading of the signal, and can be approximated by $\tau = 1/(\textit{Doppler Spread})$. Values of about 20 Hz were observed for the collection period discussed here [9] giving a value of $\tau = 50$ mS. Hence for 1.6 seconds of data, this corresponds to an effective block size of $K = 32$. Although a higher value would be better from the accuracy standpoint, this value was not considered unreasonable.

In processing the actual data, the estimated signal azimuth and elevation spread values



(a)



(b)

Figure 21: SML estimated bearing results for high latitude off-air data collected on January 23rd 1996. The results are shown for (a) azimuth, and (b) elevation.

were normally found to be less than 60° except for signals with low SNR. Given the likelihood that the very large spread estimates (> 60°) were errors due to noise, the spread values used by the SML algorithm were limited to the range from 0° to 60° to avoid unnecessarily increasing the computational requirements.

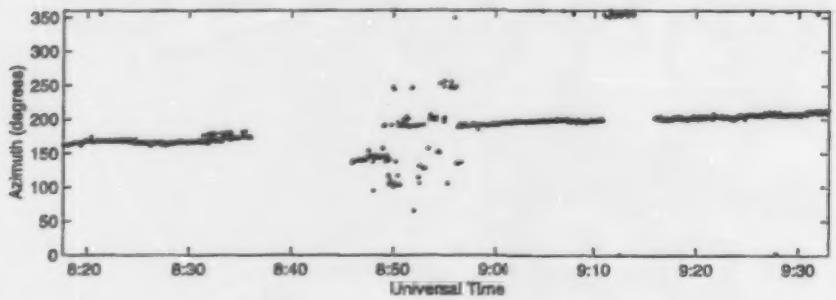
The results of processing the data using the SML algorithm are shown in Figure 21. Since the number of signals was unknown in this case, the data was processed assuming $M = 3$ signals while also retaining the results for the intermediate cases of $M = 0$, $M = 1$, and $M = 2$. Bearing estimates for data block k were then compared to data blocks $k - 1$ and $k + 1$ for the $M = 3$ case. Real signals were counted as those whose estimates varied by less than one beamwidth over the span of the three blocks. Additionally, estimated signal bearings were only counted if their corresponding estimated SNR exceeded -5 dB. Estimates not meeting both criteria were assumed to be due to noise or other error mechanisms. The results for a given data block were then taken according to the case $M = m$ where m was the number of real signals counted.

The results for the SML algorithm are considerably more informative than those of the MUSIC algorithm, with a number of different "tracks" evident. The track corresponding to the signal with most power (dominant) is shown in Figure 22 which is similar to the results in Figure 20 except that there is less scatter in the estimates. The estimated signal spread for the dominant signal is shown in Figure 23. Interestingly, given the observed trend in the azimuth spread, the lowest spread occurs during the period 8:48 to 8:52 UT when the azimuth signal direction corresponded to the great circle bearing.

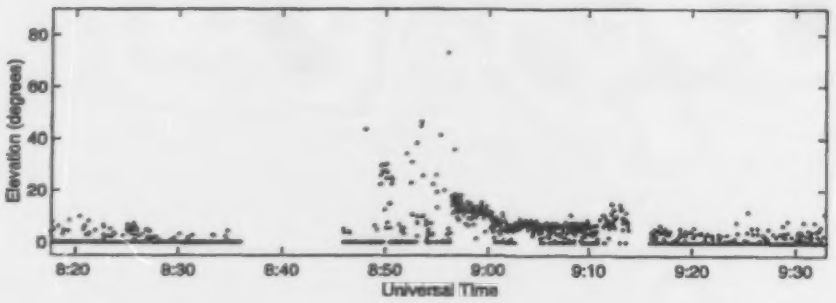
The estimated power of the dominant signal and of the noise are shown in Figure 24. The drop-outs in the signal power results (the measurements where signal power dropped to approximately 0 dB after being at higher levels in previous measurements) correspond to measurements when the transmitter was off for all or part of the measurement period. Inspecting the estimated noise power results reveals that modeling of the data was less than ideal as the noise power was correlated with the signal power resulting in noise power estimates up to 10 dB higher than the true power level (which could be measured during periods when the transmitter was off). The cause of the modeling error is probably due to array calibration problems. For example, mutual coupling will have had an effect on the antennas selected from the inner circular ring of the Pusher array, but this was not accounted for in the processing. This kind of miscalibration would then "warp" the apparent spatial distribution of a received spread signal causing errors.

The other signal tracks shown in Figure 21 were due to signals with low SNR (typically

less than 0 dB) which leaves open the possibility that some of these tracks may have been processing artifacts due to the calibration problems already mentioned. Despite these problems, however, the SML algorithm still produced results which had much less scatter than MUSIC. Additionally, the parameter estimates from the SML algorithm could be more easily used to determine the number of signals by comparing sequential signal parameter estimates, to filter out wild bearings using SNR, and to provide a useful quality estimate based on both SNR and signal spreading.

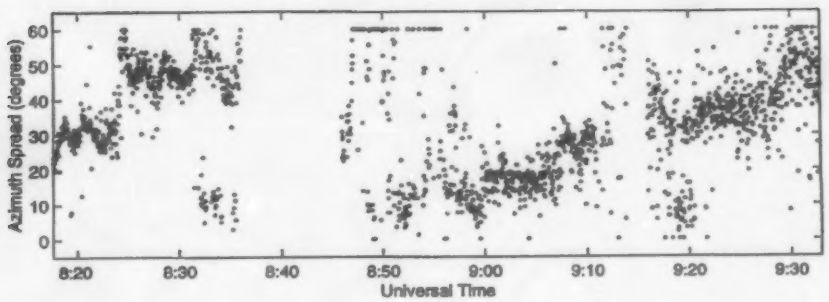


(a)

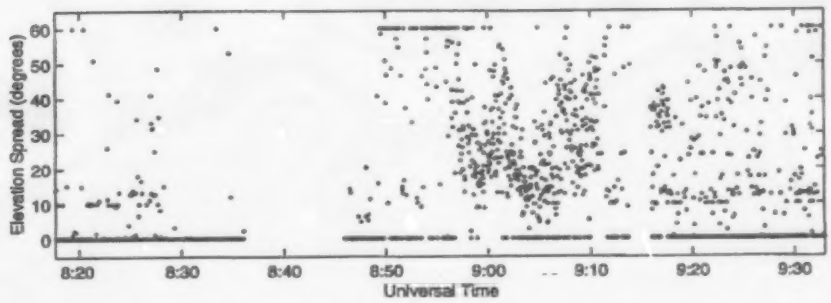


(b)

Figure 22: SML results for high latitude off-air data collected on January 23rd 1996. Only the estimated bearing corresponding to the dominant signal is shown for (a) azimuth, and (b) elevation.

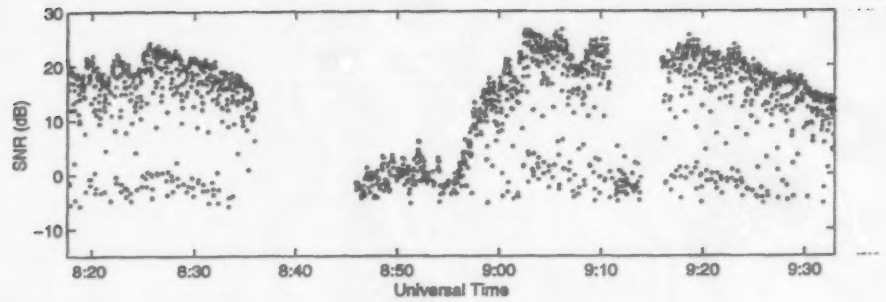


(a)

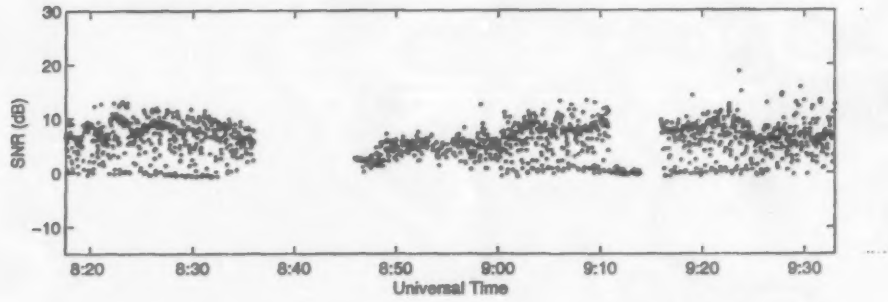


(b)

Figure 23: SML spreading results for the dominant signal showing the (a) estimated azimuth spreading, and (b) the estimated elevation spreading. Spreading estimates were limited to the range from 0° to 60°.



(a)



(b)

Figure 24: SML results showing (a) the estimated power of the dominant signal and (b) the estimated noise power. Both power levels are shown relative to the true noise power.

8.0 CONCLUSIONS AND RECOMMENDATIONS

In this report a new direction finding estimator based on maximum likelihood principles was introduced. It is distinguished from other maximum likelihood DF estimators by the fact that it deals with signals that are spatially spread in terms of both the received azimuth and elevation angles. For this reason it is called the Spread Maximum Likelihood (SML) algorithm.

The implementation of the actual algorithm relies loosely on an approach called the alternating projection maximization method [11]. The main differences are that it performs more checking of intermediate solutions (for improved accuracy) and it uses a gradient ascent technique as the search engine. The gradient ascent technique has also been modified from the normal approach to improve convergence speed.

Testing of the SML algorithm to provide guidelines on the various control parameters was performed through simulation. These parameters included the grid spacing used for modeling the spread signals, the data block size (the amount of data required to generate the data covariance matrix), the spread signal shape, and the assumed number of signals. The testing showed that these parameters could be varied for a range of values over which the performance of the SML algorithm was either almost unaffected, or was very predictable.

Testing of the SML algorithm was also carried out through simulation to evaluate its performance as a function of various conditions of the signal environment. This included the effect of noise, signal spreading, and the detection of a weaker signal in the presence of a stronger signal. In all cases, the performance of the SML algorithm was predictable, and as good as or better (particularly for spatially spread signals) than the MUSIC algorithm (a modern superresolution DF algorithm).

Finally, testing was performed using off-air data collected at CFS Alert which is located in the Arctic. Again the results showed that the SML algorithm outperformed MUSIC in terms of the variance in the bearing estimates. The SML estimates also included the amount of signal spreading (MUSIC does not) which was on the order of tens of degrees in azimuth. From an understanding of the mechanisms giving rise to signal spreading, these measurements alone indicate that the signals were scattered from moving patches. The observed trend of the bearing estimates over time essentially confirmed this. Although no evidence of sporadic-E propagation was found for this particular data set, it was observed

that the lowest amount of signal spreading occurred when the estimated azimuth bearing coincided with the great circle bearing.

The main advantage of the SML algorithm is that, compared to conventional and other modern superresolution techniques, it better models the high latitude HF signal environment. This leads to more accurate bearing estimates, less confusion in the interpretation of the results (MUSIC interprets a spread signal as a multitude of point-source signals), and a greater ability to detect a weaker point-source signal in the presence of stronger spatially spread signals. This last point is important since a point-source signal will generally yield a bearing estimate close to the great circle bearing while a spread signal will not. Additionally, the ability of the SML algorithm to measure spatial signal spreading provides valuable information on the reliability of the corresponding bearing estimate.

The main disadvantage of the SML algorithm is that it is computationally very intensive which makes it an order of magnitude or more slower than other approaches. For more benign propagation conditions, current superresolution techniques such as MUSIC might be more appropriate for use.

There are two main recommendations stemming from the work detailed in this report. The first recommendation is that for high latitude operational HF DF sites, given its benefits, the SML algorithm should be employed in some form. For example, given the computational requirements of the SML algorithm, this algorithm could be used in conjunction with faster approaches such as MUSIC. The faster approach could be used until spread conditions are suspected, such as when a number of signal bearings are estimated from similar directions, at which point, the SML algorithm would then be used. The second recommendation is that future research should focus on developing realtime implementations of the SML algorithm. Coupled with the ever increasing processing speed of modern computers, a realtime implementation could be a practical reality within a few years.

REFERENCES

[1] Dumas, D.J., "High-Latitude HF Direction Finding: A Case Study and Modeling Results", Second Symposium on Radiolocation and Direction Finding, Southwest Research Institute, San Antonio, Texas, November 1997.

[2] To, A., Wong, W., and Meng, Y.,"Direction Finding of Ionospheric Scattering Sources for Vortex Antenna System", Defence Research Establishment Ottawa, Contract Report W7714-5-99925, March 1997.

[3] Marple, S.L., *Digital Spectral Analysis with Applications*, Alan Oppenheim, editor, Prentice-Hall, Englewood Cliffs, New Jersey, Chapter 4, 1987.

[4] Chan, A.J., "Application of High-Resolution Techniques to the Joint AEDOA Estimation with a Uniform Circular Array", Master Thesis, McMaster University, Hamilton, December 1991.

[5] Stoica, P., Sharman, K.C., "Maximum Likelihood Methods for Direction-of-Arrival Estimation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 38, No. 7, July 1990.

[6] Bohme, J.F., "Source-Parameter Estimation by Approximate Maximum Likelihood and Nonlinear Regression", *IEEE Journal of Oceanic Engineering*, Vol. 10, No. 3, July 1985.

[7] Hung, E.K., Hu, Z., "Analysis of the HF Data Measured at Leitrim, Using a Matrix Calibration Method Developed at DREO", Defence Research Establishment Ottawa, Tech. Report DREO-1319, November 1997.

[8] Jenkins, R.W., "Preliminary Analysis of Kestrel Data", Department of Communications/Communications Research Centre, Tech. Memo DRL/TM083/92, February 1992.

[9] Dumas, D.J., "Direction Finding Algorithms and Software for a Sampled-Aperture Antenna Array Operating at High Latitudes in the HF Band", Industry Canada/Communications Research Centre, Tech. Memo VPRB 03/98, March 1998.

[10] Jenkins, R.W., "A Field-Aligned Scatttering Model for High-Latitude Propagation", Department of Communications/Communications Research Centre, Tech. Memo DRL/TM095/92, December 1992.

REFERENCES

[11] Ziskind, I., and Wax, M., "Maximum Likelihood Localization of Multiple Sources by Alternating Projection", *IEEE Transactions on Acoustices, Speech, and Signal Processing*, Vol. 36, No. 10, October 1988.

[12] Jenkins, R.W., "Effects of Antenna Array Geometry and Element Pattern Uncertainty on High-Latitude HF Direction Finding", Communications Research Centre, Report No. 97-006, December 1997.

[13] Jenkins, R.W., "Comparison of Spread and Point-Source Multiple-Direction Estimation Techniques for High Latitude HF Direction-Finding", Communications Research Centre, Technical Note No. 98-002, April 1998.

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. **ORIGINATOR** (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
DEPARTMENT OF NATIONAL DEFENCE
OTTAWA, ONTARIO K1A 0Z4

2. **SECURITY CLASSIFICATION** (overall security classification of the document including special warning terms if applicable)

UNCLASSIFIED

3. **TITLE** (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

A MAXIMUM LIKELIHOOD HF DIRECTION FINDING ESTIMATOR FOR HIGH LATITUDE DISTRIBUTED SIGNALS (U)

4. **AUTHORS** (Last name, first name, middle initial)

READ, WILLIAM, J.L.

5. **DATE OF PUBLICATION** (month and year of publication of document)	6a. **NO. OF PAGES** (total containing information. Include Annexes, Appendices, etc.)	6b. **NO. OF REFS** (total cited in document)
APRIL 1999	83	13

7. **DESCRIPTIVE NOTES** (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.

DREO REPORT

8. **SPONSORING ACTIVITY** (the name of the department project office or laboratory sponsoring the research and development. Include the address.

Defence Research Establishment Ottawa
Department of National Defence
Ottawa, Ontario Canada K1A 0Z4

9a. **PROJECT OR GRANT NO.** (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	9b. **CONTRACT NO.** (if appropriate, the applicable number under which the document was written)
5BB15	

10a. **ORIGINATOR'S DOCUMENT NUMBER** (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. **OTHER DOCUMENT NOS.** (Any other numbers which may be assigned this document either by the originator or by the sponsor)
DREO TR 1999-086	

11. **DOCUMENT AVAILABILITY** (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved, releasable to NATO, TTCP countries only
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify)

12. **DOCUMENT ANNOUNCEMENT** (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). however, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM

/

F2

13. **ABSTRACT** (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

(U) This report details the development and evaluation of a new direction finding estimator, the Spread Maximum Likelihood algorithm, targeted for high latitude HF propagation conditions. Of particular concern are patches of enhanced electronic density and associated instabilities in the ionosphere which drift across the polar cap during darkness causing scattering of the propagated signal over a range of azimuth and elevation directions (i.e. the signals are spatially spread). The new estimator incorporates a spatial spreading model allowing it to simultaneously deal with both patch scattered signals and signals propagated by more normal propagation mechanisms, as well as to distinguish between the two types. Evaluation of the new estimator using both simulation and collected data show it to be considerably superior to both conventional and modern superresolution approaches for high latitude propagation conditions.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

HF DIRECTION FINDING
MAXIMUM LIKELIHOOD EXTIMATION
MUSIC
N-CHANNEL
ANTENNA ARRAYS
HIGH LATITUDE HF RADIO PROPAGATION